Synthetic yeast based cell factories for vanillin-glucoside production

Strucko, Tomas

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Synthetic yeast based cell factories for vanillin-glucoside production

Tomas Strucko

PhD Thesis

Supervisors

Prof. Uffe H Mortensen (DTU Systems Biology)

C.S.O. Jørgen Hansen (Evolva A/S)

Department of Systems Biology
Technical University of Denmark
December, 2013

Preface and acknowledgments

This thesis presents the main results obtained during my PhD project, which was carried out from July 2010 to December 2013 at Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark. This work was funded by a grant from The Danish Council for Strategic Research (DSF) in Health, Food and Welfare (Grant: 09-067059).

Foremost, I would like to thank my supervisor Prof. Uffe H Mortensen for excellent guidance and support. His passion for research and enthusiasm has made my project great and, more importantly, a challenging journey in an exciting world of science. I would also like to thank my co-supervisor Dr. Jørgen Hansen, for constructive and helpful discussions and suggestions concerning vanillin related projects. I am also grateful to Dr. Kiran R. Patil (EMBL, Heidelberg), without whom this PhD project wouldn't have started for me.

During my PhD project, I was lucky to meet and work with many nice people. I am grateful to all former and present colleagues at our department for their support help and assistance. Specially, I would like to thank Line D Buron, for teaching me how to handle yeast and for being a great companion in laboratory. I owe big thanks to Zofia D Jarczynska and Christina S Nødvig for their involvement and big help in a gene amplification project. I thank Jette Thykær and Anna E Lantz for their help in planning my fermentation experiments, Anna-Lena Heins and Ted Johanson for teaching me how to run these experiments. Olivera Magdenoska, your help with intracellular metabolite measurements is highly appreciated. My work would not be possible without technical assistance from Martin Kogle, Lotte I Jarlnæs, Martin Nielsen, Tina Johansen, Jette Mortensen, Hanne Jacobsen and Elisabeth Krøger. Mhairi Workman, thank you for your support and help fighting my broken English. My students, Jens C Eriksen, Vladimir Radivojsa and Kathrine Lassen deserve thanks for their enthusiasms, hard work and dedication. My office mates Martin Schalén and Gustav H. Hansen, thank you for good working and social atmosphere in our office! Ana R Brochado, Aleksej Zelezniak, Ana Rems, Anne M Lund, Alexander Rosenkjær, Stig Rattleff, Lars Poulsen, Louise Mølgaard, Dorthe K Holm , Jakob B Nielsen, Peter B Knudsen and Paiman Khorsand-Jamal thank you for all good times we have had! Lastly, I am thankful to our administrative personnel especially to Anne N Johansen and Birgitte KH Karsbøl for invaluable help when dealing with all "formalities".

Furthermore, I owe great thanks for external collaborators. Current and former members of Center for Biosustainability, Irina Borodina, Jerome Maury and Niels B. Jensen thank you for a nice collaboration and for helping me with FACS analysis. Big thanks goes to Esben H. Hansen from

Evolva Biotech A/S for his fruitful discussions concerning vanillin projects and for useful feedback on some parts of my thesis.

I am thankful for my old friends form Lithuania for remembering about me, and for my new friends I found in Denmark for all social activities here in Copenhagen!

Last but not least, I would like to express my deepest gratitude to my family, who remotely supported me through all these years – Ačiū! Finally, a very special thanks is reserved to my girlfriend, Lili, for her understanding and patience, especially during thesis writing period.

Summary

The yeast *Saccharomyces cerevisiae* is well a characterized microorganism and widely used as eukaryotic model organism as well as a key cell factory for bioproduction of various products. The latter comprise a large variety of scientifically and industrially relevant products such as low-value bulk chemicals and biofuels, food additives, high-value chemicals and recombinant proteins. Despite the recent achievements in the fields of systems biology and metabolic engineering together with availability of broad genetic engineering toolbox, the full potential of *S. cerevisiae* as a cell factory is not yet exploited. This will require additional insights into functionality of the production system and improved genetic engineering strategies for efficient cell factory design. The aim of this project was to develop novel genetic engineering tools that allow for rapid and efficient assembly of metabolic pathways and controlled expression/overexpression of genes of interest. *De novo* biosynthetic pathway for vanillin-β-glucoside production was employed as a model system for several case studies in this project.

In order to construct yeast cell factories fulfilling current demands of industrial biotechnology, methods allowing for the introduction of large and complex metabolic pathways need to be added to the existing repertoire. To reduce the number of gene engineering steps required for cell factory construction, a new set of integrative "EasyClone" vectors have been developed in this study. This platform enables simultaneous integration of multiple genes with an option of recycling selection markers. Moreover, EasyClone vectors combine the advantage of efficient uracil-excision reaction based cloning that allows integration of one or two genes per plasmid and *Cre-LoxP* mediated marker recycling system. As a proof of concept, it was demonstrated that using EasyClone system it is possible to simultaneously integrate three DNA fragments carrying genes encoding for either yellow, cyan or red fluorescent proteins. In addition, all genetic markers were successfully removed using Cre-mediated recombination without compromising production levels of all three fluorescent proteins.

Assembly of multi-enzyme pathways into yeast does not guarantee high production levels *per se*. Moreover, pathway engineering requires precise control over the genes of interest. In this work, a novel gene amplification system was designed for fast, controlled and efficient gene overexpression in a manner that is based on targeted integration of multiple gene copies into defined loci in the yeast genome. For a proof of concept two genes encoding red and cyan fluorescent proteins were successfully amplified up to ten copies using the developed method. Linear correlation between gene copy number and mean fluorescence intensity for both reporter proteins was observed. The system was compared to multi-copy plasmids based systems and parameters such as expression

stability and homogeneity were assessed. Moreover, the gene amplification method was further applied for balancing vanillin-β-glucoside production in *S. cerevisiae.* It was previously demonstrated that *de novo* biosynthetic pathway is not capable to efficiently convert its precursor metabolite into vanillin-β-glucoside, which resulted in significant accumulation of several intermediates. Here, the gene amplification system was used to systematically overexpress individual genes or gene combinations of the biosynthetic pathway. Using this strategy, metabolic bottlenecks were identified and the production yield of vanillin-β-glucoside was 6-fold improved.

Several *S. cerevisiae* strains are commonly used by the yeast community. Among those, the S288c and CEN.PK strain backgrounds have been most frequently applied for metabolic engineering experiments. As a result, these strains have been subjected to extensive comparison studies with respect to genotype and phenotype differences. In this study, it was investigated how strain genetic background affects heterologous production of a given product. For that reason vanillin-β-glucoside biosynthetic pathway was identically reconstructed in S288c and CEN.PK strains. Comparison of two producer strains revealed that genetic background has a large impact on the vanillin-β-glucoside yield.

In summary, this work contributes with novel insights, genetic engineering tools and methodologies for improved yeast cell factory construction and metabolic engineering strategies.

.

Dansk resumé

Gæren *Saccharomyces cerevisiae* er en velkarakteriseret mikroorganisme, der i vid udstrækning anvendes som eukaryot model organisme, og som en central celle fabrik til bioproduktion af forskellige produkter. Sidstnævnte omfatter et stort udvalg af videnskabeligt og industrielt relevante produkter såsom lav værdi bulk- kemikalier og biobrændstoffer, tilsætningsstoffer og høj værdi kemikalier og rekombinante proteiner. Trods de seneste resultater inden for systembiologi og metabolic engineering, sammen med tilgængeligheden til en bred genteknologiskværktøjskasse, er det fulde potentiale af *S. cerevisiae* som en cellefabrik, endnu ikke udnyttet. Dette vil kræve yderligere indsigt i funktionalitet af det produktive system og forbedrede genetiske strategier til effektive cellefabrik design. Formålet med dette projekt var at udvikle nye genteknologiske værktøjer, der giver mulighed for hurtig og effektiv samling af metaboliske veje og kontrolleret udtryk af gener af interesse. *De novo* biosyntesevejen for vanillin-β-glucosid produktion blev fastsat som model for en række eksampler i dette projekt.

For at konstruere cellefabrikker i gær, der opfylder de nuværende industrielle bioteknologiske krav, nye metoder som giver mulighed for indførelse af store og komplekse metaboliske veje skal føjes til det eksisterende repertoire. For at reducere antallet af gen-engineering trin, der kræves for cellefabrik konstruering, er et nyt sæt af integrative "EasyClone" vektorer blevet udviklet i forbindelse med dette projekt. EasyClone giver mulighed for simultan integration af flere gener med mulighed for genbrug af selektionsmarkører. Desuden kombinerer EasyClone vektorer fordelen ved effektiv uracil - excision reaktion baseret kloning, der tillader integration af en eller to gener pr. plasmid, med det *Cre-LoxP* medieret markør genbrugssystem. Som et proof of concept, blev det påvist at ved brug af EasyClone systemet, er det muligt samtidig at integrere tre DNA-fragmenter, der bærer gener kodende for enten gul, cyan eller røde fluorescerende proteiner. Desuden blev alle genetiske markører fjernet under anvendelse af Cre-medieret rekombination uden at kompromittere produktionen af alle tre fluorescerende proteiner.

Montering af multi-enzymveje i gær garanterer i sig selv ikke et højt produktionsniveau. Desuden kræver pathway engineering præcis kontrol over generne af interesse. I dette arbejde blev et nyt genamplifikationssystem designet til hurtig, kontrolleret og effektiv genoverekspression. Systemet blev baseret på målrettet integration af flere genkopier ind i et defineret loci i gærgenomet. Som proof of concept blev to gener, der koder for røde og cyan fluorescerende proteiner forstærket op til ti kopier ved hjælp af den udviklede metode. Lineær korrelation mellem genkopitallet og gennemsnitlig fluorescens for begge reporter -proteiners intensitet blev iagttaget. Systemet blev sammenlignet med multi-kopi plasmid baserede systemer og parametre udtryk-stabilitet og

homogenitet blev vurderet. Desuden blev denne genamplifikation metode yderligere anvendt til balancering af vanillin- β-glucosid produktion i *S. cerevisiae*. Det er tidligere blevet vist, at de novo biosyntesevejen ikke er i stand til effektivt at omdanne forløber metabolitten til vanillin-β-glucosid, hvilket resulterede i signifikant akkumulering af flere mellemprodukter. I dette arbejde blev genamplifikationssystemet anvendt til systematisk at overudtrykke enkelte gener eller gen-kombinationer af biosyntesevejen. Under anvendelse af denne strategi blev de metaboliske flaskehalse identificeret og produktionsudbyttet af vanillin-β-glucosid blev forøget 6 gange.

Adskillige *S. cerevisiae*-stammer er almindeligt anvendt i gærforskningssamfundet. Blandt disse har baggrundsstammerne S288c og CEN.PK været de mest anvendte i metabolic engineering eksperimenter. Som et resultat af dette, har disse stammer været underkastet omfattende sammenligninger med hensyn til genotype og fænotype forskelle.I dette studie blev det undersøgt, hvordan stammens genetiske baggrund påvirker heterolog produktion af et givet produkt. Derfor blev vanillin-β-glucosid biosyntesevejen rekonstrueret identisk i S288C og CEN.PK stammerne. Sammenligning af to produktions stammer viste, at den genetiske baggrund havde en stor indflydelse på vanillin-β-glucosid udbyttet.

Sammenfattende bidrager dette arbejde med nye indsigter, genetisk engineering værktøjer og metoder til forbedret konstruktion af gærcellefabrikker og metabolic engineering strategier.

List of Abbreviations

3DSD	3-dehydroshikimate dehydratasev (*Podospora pausiceta*)
ACAR	Aromatic carboxylic acid reductase (*Nocardia sp.*)
ADH6	Alcohol dehydrogenase 6
BGL1	1,3-exo-glucanase 1
CDS	Coding DNA Sequence
Cre	Cre- Recombinase
DW	Dry weight
EntD	Panthen phospah transferase (*Escherichia coli*)
GA	Gene Amplification
GOI	Gene of Interest
GRAS	Generally Recognized as Safe
HsOMT	O-methyl transferase (*Homo sapiens*)
IVAN	Isovanillin
KanMX	G418 resistance cassette
OD_{600}	Optical density at wavelength (λ=600nm)
ORF	Open reading frame
PAC	Protocatechuic acid
PAL	Protocatechuic aldehyde
PCR	Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
UGT72E2	1 UDP- glycosyltransferase (*Arabidopsis thaliana*)
USER	uracil-specific excision reaction
VAC	Vanillic acid
VAL	Vanillin alcohol
VALG	Vanillyl alcohol glucoside
VAN	Vanillin
VG	Vanillin-β-glucoside
WT	Wilde type
YFG	Your Favorite Gene

Table of Contents

Table of figures

CHAPTER 5

CHAPTER 6

Chapter 1

Introduction and Outline

Nature offers a great spectrum of chemical variety with extremely diverse properties and broad fields of application, resembling food additives, cosmetics, pharmaceuticals and *etc*. Several ways can be (or are) used to acquire these products; i.e., by extraction from natural sources or by chemical synthesis, using crude starting materials. However, isolation of structurally complex compounds from their natural habitats in most cases is low yielding and time consuming, and chemical synthesis of these products is often extremely difficult or not possible [1]. Alternatively, these products can be produced by employing the bio-synthetic capacity of microorganisms.

Microbial fermentations have a long association with human activities such as production of alcoholic beverages and fermented foods. As early as in the 1920s the first large-scale microbial production of citric acid by filamentous fungi was introduced. The success of the latter paved the way for industrial scale microbial synthesis of range of other products (the most notable example - penicillin). However, the variety of early products was restricted to metabolites that were naturally synthesized (and in sufficient amounts) by the microorganisms applied. Introduction of the genetic engineering tools in the 1970s opened the possibility for production of the compounds which were not native to the microbial hosts which led to the establishment of what today is called biotechnological industry. The advent of metabolic engineering [2], [3], synthetic biology [4] and systems biology [5] approaches as the tools for engineering of microbes with enhanced and desirable properties has further accelerated development on industrial biotechnology. Today, biosynthesis of desired products is usually achieved by transferring genes or entire metabolic pathways from rare or genetically intractable organisms to well-characterized, safe and industrially relevant hosts also called “microbial cell factories”.

The yeast *Saccharomyces cerevisiae* is an attractive production host which is rigorously explored and already used as a platform for biofuel production, commodity chemical biosynthesis and production of pharmaceuticals [6]–[8]. In order to construct a yeast cell factory that would fulfill the current demand of industrial biotechnology, large and complex metabolic pathways have to be manipulated which involves multiple rounds of genetic engineering. Despite a broadly available metabolic engineering toolbox, strain construction and optimization steps are still time consuming and represent major bottlenecks in development of efficient and cost competitive production hosts

(see **Figure 1.1**). Therefore, in addition to the well-established genetic engineering tools for yeast, there is a need for novel technologies that can be used for improved yeast strain construction and metabolic engineering strategies.

The aim of this study

The potential of *S. cerevisiae* as a cell factory is far from being fully exploited. This will require additional insights into how the production system works and additional genetic tools for better or faster cell factory design. The main aim of my project was therefore to expand the genetic engineering toolbox that allows for rapid construction of cell factories as that can be used to provide insights into the functionality of the production system. In this work several key points in yeast cell factory engineering were addressed, these included i) assembly of a heterologous pathway, ii) control of heterologous gene expression levels and iii) choice of the optimal yeast strain background for heterologous production of a given product. Therefore the main goals for this project were (also see **Figure 1.1**):

i. To improve an available gene integration platform for efficient transfer of complex multi-gene metabolic pathways into yeast genome (**Chapter 3**).
ii. To develop and test a novel system for stable gene amplification that allows overexpressing gene/s of interest in a controlled manner for improved yield and identification of bottlenecks in a defined pathway (**Chapters 4** and **5**).
iii. To benchmark production capacities of two commonly used *S. cerevisiae* strains for heterologous production of vanillin-β-glucoside (**Chapter 6**).

To achieve the aforementioned goals two commonly used *S. cerevisiae* laboratory strain backgrounds S288c and CEN.PK were chosen as platform cell factories and were also used for benchmarking experiments towards heterologous production. The industrially relevant *de novo* vanillin-β-glucoside biosynthetic pathway was employed as a model system for several case studies in this project (**Chapters 5** and **6**).

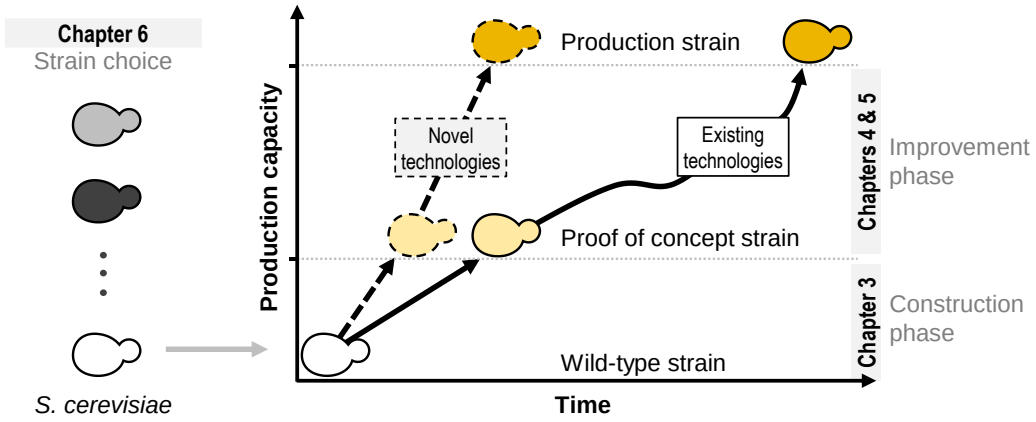


Figure 1.1. The development of a yeast cell factory for production of a given product. Production capacity refers to yields, titer or productivity. With currently available metabolic engineering tools it usually takes three to six years to improve a proof of concept strain to the final industrially compatible strain [8].

Outline

The thesis contains a literature overview (**Chapter 2**) that will introduce concepts of microbial cell factories and metabolic engineering of yeast. This chapter is followed by descriptions of the experimental work and results that I have obtained during my PhD project (**Chapters 3 – 6**). These chapters can be read independently, but it is recommended to understand the technique explained in the **Chapter 4** before reading Chapter **5**. Lastly, **Chapter 7** concludes the thesis and provides future perspectives. The content of these chapters are briefly outlined below.

Chapter 2 provides with a general background to microbial cell factories and their role in modern industrial biotechnology with the emphasis to yeast *Saccharomyces cerevisiae* as a production host. Moreover, the available biological toolbox used for design and construction of yeast based cell factories is discussed.

Chapter 3 describes the construction and validation of a new set of "EasyClone" vectors, which can serve as a platform for fast, efficient, and stable multiple gene integration into *S. cerevisiae* genome. Development of yeast strains for production of chemicals and pharmaceuticals requires iterative rounds of genetic engineering. The EasyClone set was designed in a way that allows for simultaneous genomic integration of multiple genes with a marker recycling option. In this study, a vector set was tested by simultaneous integration of genes encoding cyan, yellow and red fluorescent proteins (CFP, YFP and RFP, respectively) into separate vectors and analyzing for co-expression of proteins by flow cytometry.

In **Chapter 4,** a novel Gene Amplification (GA) system in *S. cerevisiae* is presented. When building cell factories, balanced gene expression levels play a major role in successful production of a given metabolite. Generally, biosynthesis is hampered by low activity of heterologous enzymes in the host cell. The GA was designed for fast, controlled and efficient gene amplification in a manner that is based on targeted integration of multiple copies into defined loci in the yeast genome by a system inspired by the mechanism of homing endonuclease gene transfer. As a proof of concept, overexpression of genes encoding for red and cyan fluorescent proteins (RFP and CFP, respectively) was performed. Moreover, the system was compared to multi-copy plasmids based systems and parameters such as expression stability and homogeneity were assessed.

Chapter 5 is a continuation of the work presented in **Chapter 4** and it demonstrates one of many possible applications of the GA system. Previously, a multi-step synthetic pathway for production of vanillin-β-glucoside in *S. cerevisiae* has been analyzed in detail [9]. In this study it was shown that several intermediates towards vanillin-β-glucoside accumulate, as substrate entering the pathway cannot efficiently be converted into the final product. To eliminate possible metabolic bottlenecks, I therefore used GA system as a tool to improve production of heterologous vanillin-β-glucoside by overexpressing individual genes or gene combinations of the biosynthetic pathway. Using this strategy, the production yield was significantly improved.

In **Chapter 6**, I have compared two commonly used *S. cerevisiae* laboratory strains for production of vanillin-β-glucoside. Several *S. cerevisiae* strains are commonly used by the yeast community. Among those, the S288c and CEN.PK strain backgrounds have been most frequently applied for metabolic engineering experiments. By comparing production of vanillin-β-glucoside in S288c and CEN.PK, I demonstrate that genetic background has a large impact on the vanillin-β-glucoside yield.

Chapter 7 provides with conclusions and future perspectives.

NOTE TO READER! Chapters 4 to **6** are intended to be published in peer-reviewed international journals and are written in a manuscript format, therefore, minor redundancies might appear in the text of this thesis.

List of manuscripts and publications included in this thesis

CHAPTER 3

N. B. Jensen, **T. Strucko**, K. R. Kildegaard, F. David, J. Maury, U. H. Mortensen, J. Förster, J. Nielsen and I. Borodina. **EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*.** *Published: FEMS Yeast Research* DOI: 10.1111/1567-1364.12118

CHAPTER 4

Line D Buron*, **Tomas Strucko***, Christina S Nødvig, Zofia D Jarczynska, Louise Mølgaard, Uffe H Mortensen. **Controlled gene amplification enables high tunable, selection free gene expression in the yeast *Saccharomyces cerevisiae*.** *Manuscript is in preparation. *-The authors contributed equally to this work.*

CHAPTER 5

Tomas Strucko, Esben H Hansen, Jørgen Hansen, Uffe H Mortensen. **Balancing the heterologous biosynthetic pathway for improved production of Vanillin-β-glucoside in *Saccharomyces cerevisiae*.** *Manuscript is in preparation.*

CHAPTER 6

Tomas Strucko and Uffe H Mortensen. **Benchmarking two commonly used *Saccharomyces cerevisiae* strains for heterologous vanillin-β-glucoside production.** *Manuscript is in preparation.*

Chapter 2

General Background

Importance of industrial biotechnology

The concept of *biotechnology* integrates a wide range of methods and processes for exploiting and modifying of living organisms for production of useful goods and has been used by humankind for thousands of years mainly for food production and agricultural purposes. Definition of this concept has changed over the years and today it is closely associated with microbial biotechnology accompanied with genetic engineering and microbial fermentation technologies. When biotechnology is applied for purposes of different industrial sectors by use of microbes as cell factories it is commonly defined as *industrial biotechnology* (in Europe known as *white biotechnology*). The major progress of *modern* biotech industry started after discovery of genetic engineering techniques which made possible production of recombinant proteins in microbial cell factories. In 1982 human insulin produced by engineered *Escherichia coli* was the first recombinant DNA drug to appear on the market [10]. Since then several other microorganisms have been successfully adopted for microbial bioproduction of heterologous products of which pharmaceutical proteins were primary targets due to their demand and high price in the market [11].

Availability of genetic engineering tools made it also possible to synthesize chemical industry related products in microbial hosts, however not in a cost-efficient or cost-competitive manner compared to traditional methods (e.g., chemical synthesis routes using petrochemicals as raw materials). This, however, has changed dramatically within last two decades. Substantial advancements in technology and scientific areas such as multidisciplinary field of metabolic engineering and systems biology greatly expanded the palette of products that can be synthesized via microbial cells. Today it is conceivable to design and develop a cell factory for the production of virtually any relevant organic molecule [12]. The huge variety of biotechnology related products can be roughly grouped in two categories: low-value (fuels, bulk chemicals) and high-value (fine chemicals, pharmaceutics).Recent progress in production of above mentioned products via various microbial cell factories is summarized in a review [12].

One needs to add that industrial biotechnology related processes offer several advantages such as reduced pollution and energy consumption compared to classical methods i.e., chemical synthesis.

Moreover, the ability of microorganisms to utilize renewable biomass based materials offers sustainable alternatives for petrochemical based industry. Dwindling resources of fossil fuels, increase in CO_2 emissions and overall environmental concern has turned industrial biotechnology into a key technology for future development of biosustainable production [13]. It is therefore predicted that microbial based production will substitute up to 20% of products of chemical industry by year 2020 [11]. In a recent reviews [14], [15] it is also anticipated that future production of biofuels and other desired chemicals will take place in so called "biorefineries", where different types of biomass will be processed into sugars and subsequently converted into the desired products through microbial fermentation [15].

To this end, the future and success of industrial biotechnology greatly depends on a constant improvement of technologies allowing fast construction and optimization of microbial cell factories that can cost efficiently convert the sugar-based raw materials into products of interest [11]. Depending on the desired product and its further applications different types of microbial cells can be used for biosynthesis, commonly these are several variants of bacteria, yeast and filamentous fungi. The yeast *Saccharomyces cerevisiae* has a tremendous application potential in biotech industry, moreover, it is forecasted to become a key cell factory for production of biosustainable chemicals in the near future [15].

Saccharomyces cerevisiae

From beer brewing to eukaryotic model organism

Saccharomyces cerevisiae also known as baker's yeast represents one of the most studied and characterized species out of about 1000 known yeast species in a vast kingdom of Fungi. It has earned a title of "the oldest domesticated microorganism" due to its millennia-long association with human activities in wine and beer brewing as well as bread baking [16]. This microorganism however was not discovered until the 19th century after pioneering studies of Louis Pasteur and others [17]. The name of the species reflects its ability to ferment sugar and produce ethanol – word *Saccharomyces* means sugar mold and *cerevisiae* means beer [18]. Taxonomically, *S. cerevisiae* is defined as a eukaryotic single-cell (ovoid shaped 5 to 10µm in diameter) microorganism that propagates (vegetatively) by a process called budding [19]. The life cycle of *S. cerevisiae* can alternate between two phases – haploid and diploid either of which can exist in stable cultures.

Research activities focusing on *S. cerevisiae* have resulted in accumulation of comprehensive knowledge concerning genetics, physiology and biochemistry as well as genetic engineering and

fermentation technologies. Pioneering discoveries in molecular biology (recombinant DNA technology) and introduction of highly efficient transformation methods in *S. cerevisiae* [20], [21] has opened new possibilities for basic as well as applied research areas. In 1996, full genome sequence of *S. cerevisiae* became available for scientific community [22] marking the beginning of "revolutionary" era of genomics. Right after, a comprehensive genome characterization study to refine the number and location OFRs present in the *S. cerevisiae* was done via bioinformatics [23]. The latter also made it possible to study yeast gene expression profiles by microarray hybridization [24], [25]. Moreover, availability of genome sequence in combination with gene engineering tools made particular yeast a workhorse for high throughput systematic studies using various "omics" approaches [26]. Increasing interest in the particular yeast resulted in several global scale analysis; i.e., functional profiling was done by constructing single gene deletion library of nearly all annotated (approx. 6000) OFRs of *S. cerevisiae* [27], moreover, a comprehensive study of cellular protein localization was done by assembling the library of green fluorescent tagged proteins [28]. Recently, genetic interactions with regards to phenotype was studied by assembling a double gene deletion library in yeast [29], [30]. To this end, constantly increasing amounts of information resulted in an assembly of the most comprehensive yeast genome database – *Saccharomyces* genome database (SGD) [31].

Due to its substantial role in science, it is not surprising that the yeast *S. cerevisiae* has become a widely applied model organism to study fundamental biological processes such as cell aging [32], mRNA transport [33] or cell cycle [34] and other processes [18]. More importantly, as a representative of eukaryotic organism *S. cerevisiae* served as a model for studying human diseases such as cancer [35], [36], neurological disorders such as Alzheimer's and Parkinson's [18] and other important syndromes [37].

Sub-species variation

Since the isolation of the first yeast culture by Hansen in 1900s the number of isolated/developed *S. cerevisiae* strains has increased dramatically. Today a broad collection of *S. cerevisiae* strains are available for both commercial and academic purposes. However, not until appearance of high-throughput and cheap Next-Generation DNA sequencing we could discover the extraordinary variation within a single yeast species. Large-scale genome sequencing of mainly commercial *S. cerevisiae* strains revealed existence of at least four separate lineages with a considerable genetic variation between and within these lineages [16]. Detailed analysis of DNA sequences revealed extensive cross hybridization, genomic mosaicism [16] and gene copy number variation, which is most likely a result of an artificial selection for preferred traits [38].

For academic purposes several laboratory *S. cerevisiae* strains have been developed such as FL100, BY series (S288c derivative), Sigma1278b, W303 and CEN.PK series. The variety of laboratory strains was as result of diverse research activities within different disciplines such as genetics, physiology and biochemistry. Recent genome sequencing projects have covered most of the laboratory strains (reviewed in [39]), revealing their genetic variations and relation to various *S. cerevisiae* strains (**Figure 2.1**). Genome comparison analyses have shown significant differences within laboratory yeast strains, represented by different gene copy numbers, insertions, deletions, and an extensive number of single nucleotide polymorphisms affecting thousands of genes [39]. As laboratory strains are widely used for development of industrially related processes further systems-level comparison studies have been pursued in order to address the phenotype to genotype relations of these strains. In particular, differences of the two commonly used strains for cell factory development (S288c [40] and CEN.PK [41]) have been studied extensively [42]–[44]. These differences are also addressed in **Chapter 6** of this thesis.

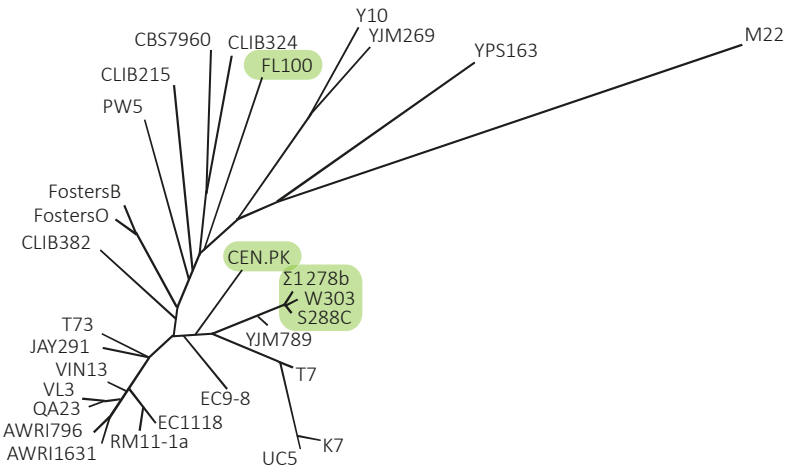


Figure 2.1. Phylogenetic three based on whole genomic distances as calculated by Nijkamp et al. [44]. The names marked in green represent laboratory strains of *S. cerevisiae*, the unmarked ones are commercial yeast strains mainly used for industrial production of beer, wine and bioethanol. The graph is modified from [39].

S. cerevisiae as a cell factory

In addition to the fact of being a well-established eukaryotic model organism [37], *S. cerevisiae* also plays a major role in applied research due to its fast growth, robustness and substantial tolerance towards a harsh industrial conditions. Without a doubt, brewing and baking represent the oldest applications of *S. cerevisiae* for industrial purposes. As a result, well-established fermentation technology and its long association with food production earned *S. cerevisiae* a GRAS (Generally Regarded as Safe) status that further increased interest in the particular yeast as a platform for bio-production. In addition, *S. cerevisiae* presents a range of advantages for industrial applications: i.e., tolerance to relatively low pH (3~5), high sugar and ethanol concentrations, thereby minimizing the risk of bacterial contamination; ability to grow in aerobic and anaerobic conditions, fast utilization of wide range of sugars; fair resistance to inhibitors found in the hydrolysates of cellulosic biomass [6].

The advent of genetic engineering tools allowed modifying the genetic landscape of yeast and transforming it into a cell factory that can produce nonnative yeast products. Since then, numerous works aimed to construct a yeast cell factory that can convert sugar-based substrates into different types of products ranging from bioethanol and low-value chemicals to therapeutic proteins and high-value chemicals (see **Figure 2.2**) [6]–[8]. **Table 2.1** lists the recent achievements/attempts in engineering of the *S. cerevisiae* for production of valuable products. Several native yeast metabolites such as succinic acid and fumaric acid can be produced as bulk chemicals. Bioproduction of biofuels is of extensive interest, in fact, *S. cerevisiae* is currently used for the world's largest industrial production of bioethanol [45]. Attempts to produce fine chemicals in yeast is highly represented by increasing number of research projects focusing on production of the particular class of compounds [6]–[8]. These are chemically diverse and structurally complex molecules, usually found in secondary metabolism of, bacteria, fungi or plants, and can be applied in food, perfume or pharmaceutical industries. An extensive list of the chemicals belong to several classes of secondary metabolites: i.e., isoprenoids (carotenoids, sesquiterpenes and diterpenoids), phenolics (flavonoids, stilbenoids), polyketides and alkaloids [46].

Most prominent among yeast cell factories producing secondary metabolites are those which synthesize precursors to the antimalarial drug – artemisinin [47]. Artemisinin is a sesquiterpene that is naturally isolated from the plant *Artemisia annua*. An alternative for natural extraction was recently proposed by Westfall and coworkers [48], here yeast strain was engineered to produce high titers of amorpha-4,11-diene. By combining microbial production of amorphadiene and developed methods for its subsequent conversion to the final product, industrially feasible

production of semisynthetic artemisinin was achieved [49]. In fact, it was estimated that the antimalarial treatments produced via particular technology will have lower than the current average prices, making it more accessible to the developing world.

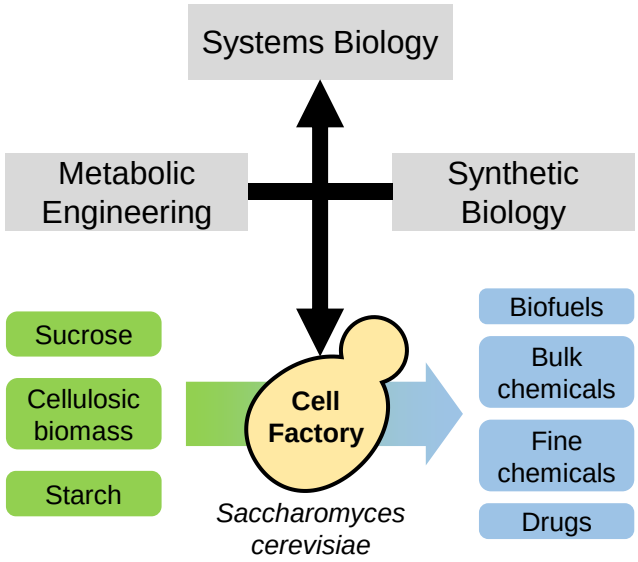


Figure 2.2. Concept of a yeast cell factory and input of various scientific disciplines to its development.

Microbial production of most of the products mentioned in this section would not be possible without multiple rounds of genetic manipulations, phenotype characterization and strain optimization. This is done through metabolic engineering strategies which are discussed in the following section.

Table 2.1. Examples of the products produced in *S. cerevisiae.* The table is based on [8].

Category	Product	Year	Strain	Reference
Biofuels	Ethanol	2007	D452-2	[50]
		2010	CEN.PK,	[51]
		2010	YPH499	[52]
	Isobutanol	2008	BY4742	[53]
		2011	CEN.PK	[54]
	Bisabolene	2011	BY4742	[55]
	Biodiesels	2012	YPH499	[56]
		2013	YS58	[57]
Bulk chemicals	Glycerol	2002	CEN.PK	[58]
	Pyruvic acid	2004	CEN.PK	[59]
	Polyhydroxyalkanoates (PHA)	2006	BY4743	[60]
	D-ribose and ribitol	2010	CEN.PK	[61]
	Succinic acid	2010	AH22	[62]
	L-lactic acid	2011	CEN.PK	[63]
	Polyhydroxybutyrate (PHB)	2013	CEN.PK	[64]
Fine chemicals	Resveratrol[1]	2003	FY23	[65]
	Farnesol[2]	2003	FL100	[66]
	L-ascorbic acid[1]	2004	W303	[67]
	Artemisinic acid[1]	2006	S288c	[47]
	β-carotene[3]	2007	CEN.PK	[68]
	Vanillin[2,3]	2009	S288c	[69]
	Non-ribosomal peptides (NRPs)[1]	2010	CEN.PK	[70]
	Casbene[1]	2010	BY4742	[71]
	Patchoulol[2]	2011	CEN.PK	[72]
	Eicosapentaenoic acid (EPA)[1]	2011	CEN.PK	[73]
	β-amyril	2011	CEN.PK	[42]
	Naringenin[1]	2012	CEN.PK	[74]
	Amorphadiene[1]	2012	CEN.PK	[48]
	Cubebol[3]	2009	CEN.PK	[75]
		2010	CEN.PK	[76]
	Vanillin-β-glucoside[2,3]	2010	S288c	[9]
		2013	S288c	[77]

[1]- pharmaceutic, [2]- fragrance, [3]- food additive

Metabolic engineering of yeast cell factories

Since the discovery of the yeast *S. cerevisiae*, there has been increasing interest to develop yeast strains with improved characteristics that could be exploited in industrial processes. Conventionally, strain improvement was done by classical genetic methods (i.e., sexual cross) or by random mutagenesis (using radiation or DNA damaging substances) followed by a subsequent selection for desired phenotype. Despite the great success, these approaches were time consuming as a high number of mutants had to be screened. Moreover, in addition to the desired traits, generated mutations were leading to accumulation of unwanted characteristics. As an alternative to the classical strain improvement methods a new concept, namely Metabolic Engineering has emerged.

Briefing on Metabolic Engineering

The concept of *metabolic engineering* was simultaneously proposed in two independent publications in 1991 [2], [3]. Bailey defined metabolic engineering as "*the improvement of cellular activities by manipulations of enzymatic, transport and regulatory functions of the cell with the use of recombinant DNA technology*" [3]. In other words, the main aim of the metabolic engineering is a rational construction of microorganisms with desired phenotypes through the targeted genetic engineering. The fact that *S. cerevisiae* is extremely well characterized and tractable to genetic modifications made this particular microorganism one of the main workhorses for industrially driven metabolic engineering applications (**Table 2.1**). Depending on the overall aim, metabolic engineering is used to increase yields of the desired products, improve metabolic fluxes through desired pathways or remove unwanted or competing reactions. In addition, the heterologous properties can be introduced by transferring foreign genes or all metabolic pathways into the yeast *S. cerevisiae*. However, introduction of heterologous pathway does not imply high levels of desired product production [78]. To solve this, several knowledge-based approaches such as local and/or global metabolic engineering can be applied. The first strategy involves manipulation of native or heterologous metabolic pathway responsible for biosynthesis of a given [47]. An example of the particular strategy is described in **Chapter 5** of this thesis. On the other hand, global metabolic engineering aims to reshape the overall metabolic network of the yeast in order to improve precursor metabolites and cofactor supply [9], up-regulate the genes involved in product export routes, or redesign native (competing) metabolic pathways so that competing reactions are alleviated [72].

Metabolic engineering is a constantly developing discipline that relies on advancing tools for genetic modifications and systems-level analytical and computational techniques (**Figure 2.2** and **Table**

2.2). In general construction of novel cell factories is done in an iterative manner following the three principles of metabolic engineering: *design*, *construction* and *analysis*. Each of these steps have a specific function during cell factory development process (reviewed in [7]). A brief explanation of each step in metabolic engineering cycle is described below.

For the *design* step, several key decision points have to be considered: i.e., choice of the product and production host and identification of target genes to be manipulated. The *design* step also involves computational strategies such as use of genome scale models in order to predict cellular responses or identify new targets for genetic engineering [79]. With regards to the choice of production host, it has been shown that *S. cerevisiae* displays a significant sub-species variation, thus is highly likely that production yields of a given metabolite might vary depending on the cells genetic background. The latter is addressed in the **Chapter 6**.

The *construction* step implements the designed strategies and mainly relies on the genetic engineering and synthetic biology tools. Here native, heterologous or synthetic genes can be integrated, genes can be overexpressed, deleted or mutated. Moreover, the ability to control expression levels of genes involved in the biosynthetic pathway of interest is of high importance. The gene expression levels can be manipulated in several ways. The simplest method is to control gene copy number which can be achieved by using yeast shuttle vectors or by genomic integration of the gene of interest (see page 17) [80]. Alternatively, gene expression can be controlled at the transcription level (i.e., by using native yeast promoters of various strength [81], [82] or synthetic promoters [83]–[85]) or translation level (i.e., by RNA regulators [86] or codon optimization).

Lastly, in the *analysis* step the constructed cell factory is characterized using various analytical methods. Technological and computational advances resulted in a development of high-throughput analysis tools also known as “omics” technologies, i.e., genomics, transcriptomics, metabolomics and etc. [7]. These approaches made it possible to analyze the effects of genetic manipulation on the cell in a global scale and provide with better understanding of the mechanisms underlying advantageous phenotypes of the constructed cell factories.

Table 2.2. Metabolic engineering with respect to other disciplines. The following strategies are described in recent reviews [5], [7].

<table>
<tr><th colspan="2">Metabolic Engineering</th></tr>
<tr><td colspan="2">Redesign or import metabolic pathways</td></tr>
<tr><th>Systems biology</th><th>Synthetic biology</th></tr>
<tr><td>High-throughput analysis</td><td>New parts and functions</td></tr>
<tr><td>• Genomics
• Proteomics
• Transcriptomics
• Metabolomics
• Fluxomics</td><td rowspan="3">• Synthetic genes (codon optimization)
• Protein engineering
• Unnatural metabolic pathways and new regulatory circuits
• Assembling synthetic organisms</td></tr>
<tr><td>Computational tools</td></tr>
<tr><td>• Genome-scale models</td></tr>
</table>

Biosynthetic pathway engineering

Biosynthetic (metabolic) pathway can be defined as a sequence of enzymatically catalyzed reactions that convert source (**S**) metabolites into a product (**P**) [87]. Nature has provided a great array of metabolic pathways represented by diversity of existent (micro)organisms, however, these pathways often need to be reshaped or most commonly transplanted from the original organism to the well-developed production platform (in this case *S. cerevisiae*). In general, engineering of metabolic pathways in yeast (as well as in other microbial hosts) can be done in several ways (**Figure 2.3**). In some cases, the host's native biosynthetic pathway can be re-designed in order to eliminate, bypass or to direct fluxes towards a given product metabolite. This is typically done by deleting, overexpressing or mutating genes of the metabolic pathway of interest. In other cases, a fraction or a whole natural biosynthetic pathway originating from one organism can be transferred into another more suitable production host, e.g., metabolic pathway found in plants is transferred into *S. cerevisiae.* Recent examples of successful integration of heterologous pathways into *S. cerevisiae* includes production of artemisinic acid [47] and amorphadiene [48] – precursor metabolites for well-known antimalarial drug artemisinin; or introduction of seven-step pathway from *Arabidopsis thaliana* for indolylglucosinolate production [88]. Moreover, multiple genes resembling fractions of biosynthetic pathways of several different organisms might be assembled to form a hybrid metabolic pathway in heterologous microorganism. In this manner, synthetic pathways can be constructed to produce novel and structurally non-natural molecules [89]. The last, and perhaps the most challenging design strategy, namely *de novo* pathway engineering, involves assembly of biosynthetic pathways using a combination of unrelated genes originating from discrete organisms to create novel metabolic routes towards compounds of interest. *De novo* metabolic pathways represent possibly the most promising strategy to exploit and expand synthetic capacity of

microbial systems making it possible to produce un-natural or natural chemicals for which original biosynthetic pathways are not known. One of the first examples of the latter is the *de novo* multi-step biosynthetic pathway for production of aroma compounds vanillin and vanillin-glucoside in baker's and fission yeast [69] and more recent example demonstrates production of key intermediate for flavonoids – naringenin [74]. Today, the *de novo* pathway design is much easier due to availability of numerous computational based tools (reviewed in [87]).

It is important to emphasize that all the above mentioned pathway engineering strategies can in theory be applied to expand host's capability to utilize new substrates, e.g., cellulose, starch, lactose, xylose, arabinose and *etc.* [8].

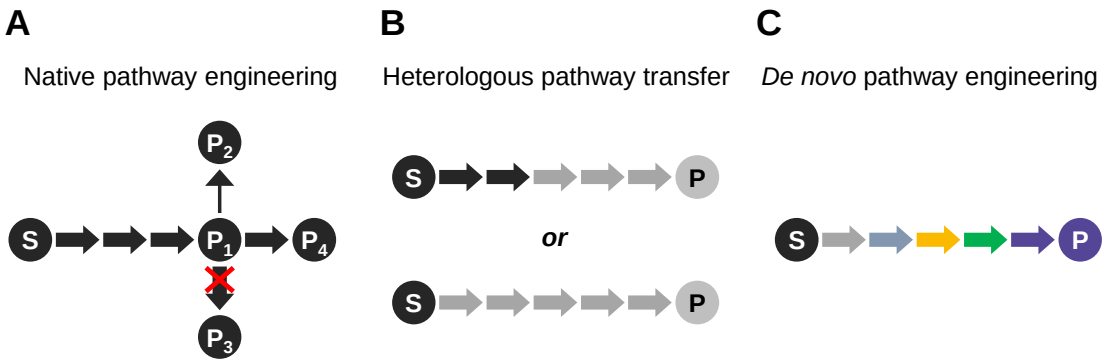


Figure 2.3. Strategies for biosynthetic pathway design.A) Parts of a native pathway can be modified in order to redirect or eliminate flux towards a given product. B) Parts or whole metabolic pathways can be transferred from one organism to another. C) Each part of a metabolic pathway is proposed individually and usually originates from different sources. S – substrate, P – product; the color of the arrows represent metabolic enzymes originating from the same organism.

Genetic engineering tools for *S. cerevisiae*

The *construction* step in metabolic engineering cycle is always based on genetic engineering. Due to extensive research efforts within yeast genetics, a vast number of genetic engineering tools have been developed. In general, a gene(s) can be introduced into *S. cerevisiae* in two ways; by using plasmid vectors or by genomic integration (**Figure 2.4**). Depending on the overall goal both methods have their advantages and disadvantages and were routinely applied for metabolic engineering purposes [80]. These methods are briefly explained in the following text.

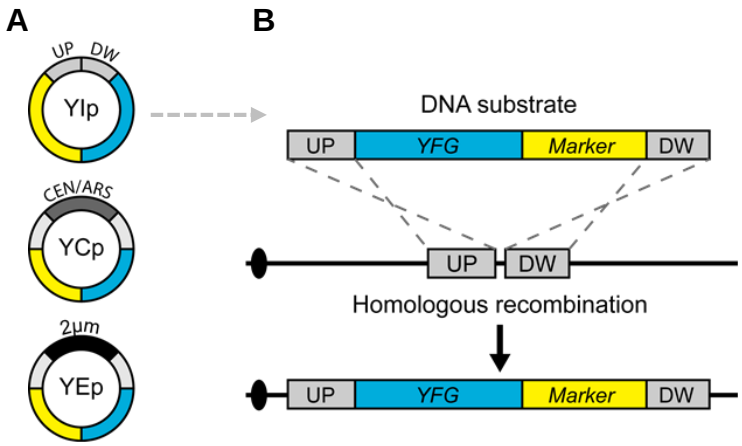


Figure 2.4. Introduction of genes into yeast *S. cerevisiae*. A) Simplified representation of different types of yeast plasmid vectors; YIp – integrative, YCp – centromeric, YEp – multicopy episomal vector. B) Genomic integration by homologous recombination. *YFG* – your favorite gene, UP and DW targeting sequences for homologous recombination.

Plasmid based expression

There are several types of extra-chromosomally replicating plasmids of which YEp and YCp are commonly used ones. Both types are *S. cerevisiae/E. coli* vectors (can propagate in both microorganisms) which contain a selection marker and typically a multi cloning site (MCS) for insertion of DNA fragments of interest. These vectors are widely used and proven effective for various applications [80]. The YCp plasmids carry an origin of replication and centromere sequence (CEN/ARS) that provides stable segregation during cell division maintaining 1-2 copies per cell [90]. Another type is YEp (or episomal) vectors, these contain a native 2µ sequence and are reported to be maintained at 10 to 40 copies per cell under selective conditions [91], [92]. The latter plasmids are usually used when high gene overexpression level is needed.

Today, an extensive YCp and YEp (and YIp – integrative plasmids discussed later) vector collection carrying various properties and modifications are available for yeast community (see **Table 2.3**). First generation plasmids were relatively simple in design and were constructed in several versions that differed in selection markers and origins of replication [93], [94]. The pRS series developed by Sikorski and Hieter [95] were constantly updated with additional selection markers [96]–[98] and were widely used for gene expression in yeast. Several new variants (mainly derivatives of pRS series) were constructed to contain a range of various constitutive and inducible promoters [99]–[101]. With increasing interest in metabolic engineering there was need to express more than one gene simultaneously, therefore, series of vectors (YCp and YEp) equipped with bidirectional promoters were assembled [81], [102]. Commercially available multicopy pESC vectors (Agilent

Technologies) with bidirectional promoters are the ones commonly used for metabolic engineering purposes [103], [104].

Table 2.3. Yeast vector series for gene expression in yeast.

Name	Type	Marker	Promoter	Ref.
YCp4XX, YEp4XX	YCp, YEp	*URA3, TRP1, LEU2, LYS2, HIS3*	—	[93]
YCplac, YEplac, YIplac	YCp, YEp, YIp,	*URA3, TRP1, LEU2*	—	[94]
pRS series	YCp, YEp	*URA3, TRP1, HIS3, LEU2, KanMX, ADE2, MET15 , hphNT1, natNT2*	—	[95]–[98]
p4XX	YIp, YCp, YEp	*URA3, TRP1, HIS3, LEU2*	*CYC1, ADH1, TEF1, GPD1, MET25, GAL1, GALS, GALL*	[99], [100]
pCMXXX series	YCp, YEp	*URA3, TRP1*	*tetO-CYC1*	[101]
pBEVY, pBEVY-G	YEp	*URA3, TRP1, LEU2, ADE2*	*TDH3-ADH1, GAL1-GAL10*	[102]
pISXXX	YIp	*URA3*	*TEF1, GAL1, STE12*	[105]
pAG (*Gateway™*)	YIp, YCp, YEp	*URA3, TRP1, HIS3, LEU2*	*TDH3, GAL1*	[106]
pY2x-GAL(1/10)-GPD	YEp	*URA3, TRP1, HIS3, LEU2*	*GAL1-TDH3, GAL10 -TDH3*	[107]
pSP-G1, pSP-G2	YEp	*URA3*	*TEF1-PGK1*	[81]
pXP series (*Cre-LoxP*)	YIp ,YCp, YEp	*URA3, TRP1, HIS3, LEU2, ADE2, MET15 (recyclable)*	*PGK1, TEF1, HXT7-391 , ADH2, GAL1, CUP1*	[108] [109]
pX, pXI, pXII series (USER™)	YIp	*URA3 (recyclable)*	*Uni- or bidirectional*	[88]
pCfB series (*Cre-LoxP,* USER™)	YIp	*URA3, HIS3, LEU2,LYS5, KanMX (recyclable)*	*Uni- or bidirectional*	[110]
Commercial vectors				
pYC, pYES	YCp, YEp	*URA3, TRP1*	*GAL1*	Invitrogen
pESC set	YEp	*URA3, TRP1, HIS3, LEU2*	*GAL1-GAL10*	Agilent

Both YCp and YEp type vectors are ideal for expression of one/two genes from a single plasmid in a fast and convenient manner. However, in order to construct a yeast cell factory for production of high value heterologous products large and complex metabolic pathways (can exceed more than ten genes) have to be expressed. Moreover, precise control of heterologous gene expression levels is crucial for balanced biosynthesis of a given metabolite. The latter would be very difficult to achieve using plasmids; first, because there is a limited number of genetic markers available, and second, it is difficult to stably maintain two or more different vectors in a single cell even if a strong selection is applied [108], [111], [112]. Non nonhomogeneous gene expression patterns when using two or three plasmids were demonstrated in the following **Chapters 3** and **4**. In addition, it has been recently shown that plasmid copy number can significantly vary depending on the promoter and selection marker used for expression of the gene of interest [113]. Chromosomal gene integration

is an alternative method for gene expression that offers precise control on copy number and segregation stability.

Chromosomal integration

Highly efficient homologous recombination machinery of *S. cerevisiae* [114] allows controlled integration of biosynthetic pathway genes. Gene integration (or deletion) can be done by transforming yeast cells with a linear DNA fragment(s) that contain homologous targeting sequences (UP and DW), gene(s) of interest and selection marker (see **Figure 2.4B**). The UP and DW sequences can be as short as 50 bp, however, routinely 200-500 bp long sequences are used to ensure a precise gene targeting. Once integrated, the DNA fragment of interest will stably segregate during cell division (assuming that no unstable structures are introduced i.e., direct or inverted repeats) without an additional selective pressure. Moreover, recyclable marker systems (e.g., *Cre-LoxP* based [115] or direct repeat recombination based systems using *URA3* or *TRP1* markers) allow iterative integrations/deletions of many genes of interest.

There are two basic methods to construct a DNA substrate for genomic integration; i.e., by using integrative vector (YIp) based or PCR based methods. YIp vectors contain the same basic structural elements as extra-chromosomal plasmids apart the replication sequence. These vectors cannot be retained in the yeast cell unless integrated. Several series of integrative vectors with the specific targeting sequences were developed over the years (**Table 2.3**). Standard YIp vectors were designed to target the auxotrophic marker regions [94], [95], [105], [106], [116]. The insertion of large metabolic pathways requires multiple targeting sites. Recently, YIp series were developed to target a number of specific genomic regions that exhibit various gene expression levels [117] or stable (high) expression levels [88].

PCR generated DNA targeting substrates have been/are widely used for gene integration and gene deletions [30], [115], [118]. PCR-based gene targeting substrates can be assembled *in vitro* by fusion-PCR [119] or *in vivo* by exploiting yeast homologous recombination [120]–[122]. To facilitate construction of PCR based integration fragments various “template plasmids” have been developed. These contain readily constructed cassettes suitable for different purposes (i.e., deletion, promoter swap or tagging) that can be PCR amplified and fused to UP and DW targeting sequences of interest. To this end, low oligonucleotide price and availability of High fidelity DNA polymerases makes, in many cases, PCR based integration methods preferably for high-throughput metabolic engineering purposes [80].

Available DNA cloning techniques

Conventionally, restriction-ligation methods were used to clone expression cassettes into above mentioned plasmids. Today, a wide array of different high-throughput cloning methods are available which over time have proven to be of great importance, e.g., Gateway™ cloning (Invitrogen) [123], Gibson Assembly® cloning (New England Biolabs) [124], Golden Gate cloning [125], Infusion cloning (Clontech) and DNA assembler [122]. For a recent review on available cloning and pathway assembling techniques refer to [126]. In this PhD project a uracil-specific excision reaction (USER)-based cloning [127] and USER fusion [128] techniques were extensively applied. USER cloning is a restriction-ligation free in vitro assembly method allowing to combine multiple DNA fragments. The technique is described in [129] and its application for yeast based systems is explained in **Chapter 3**. Moreover, the gene expression platform [88] described in the following section was designed to be compatible with USER cloning technique.

Yeast gene expression platform

As outlined previously, to fulfill the current demands of industrial biotechnology large and complex metabolic pathways have to be manipulated. Moreover, to obtain a stable yeast factory heterologous genes have to be integrated genomically in a fast and simple manner. To fulfill this, several methods allowing assembly and transfer of multi-gene metabolic pathways have been recently proposed [130], [131].

In my PhD project all yeast strains were constructed by implementing the gene integration platform (**Figure 2.5**) previously developed at our center. The platform contains a collection of custom made vectors carrying unique recombination sequences (UP) and (DW) that target designated locations on three chromosomes X, XI and XII [88]. Moreover, integration sites are grouped in clusters in order to facilitate construction of multi-gene pathways by strategies that involve sexual crossings. The main benefit of the platform is that all integration sites are separated by genetic elements that are essential for sustaining wild type growth rates (e.g., essential genes). The latter is a built-in safety system that prevents strains from propagating if they lose genes through direct repeat recombination – the spontaneous event that can occur if repetitive elements (i.e., promoters, terminators or genes) are used. Moreover, each vector in the set contains a USER cloning friendly cassette AsiSI/Nb.BsmI [132] flanked by two *S. cerevisiae* terminators *ADH1* and *CYC1*. By applying USER cloning and fusion [129] techniques it is possible to simultaneously assemble and clone one or two genes together with uni- or bidirectional promoter into a vector of choice. The success of the platform was demonstrated by reconstruction of indolylglucosinolate metabolic pathway from *Arabidopsis thaliana* [88] and *de novo* vanillin-glucoside pathway (**Chapters 5** and **6**) in to *S.*

cerevisiae. Recently, the integration sites of the particular platform were used for development of a novel Gene Amplification (GA) system described in **Chapter 4**.

Unfortunately, all integrative plasmids contain the same genetic marker which is *URA3* gene of *Kluyveromyces lactis*. This means that DNA fragments (containing one or two genes of interest) need to be integrated in an iterative manner, i.e., prior to the next round of transformation *URA3* marker has to be counter-selected. This can become time consuming when larger pathways need to be engineered. In this PhD project the integrative vectors have been updated to contain an expanded genetic marker repertoire (see **Chapter 3** or [110]). Using an updated vector set it was possible to simultaneously integrate three DNA fragments (or up to six genes) and successfully recycle all genetic markers through Cre-*LoxP* system [115] without compromising the expression levels of integrated genes.

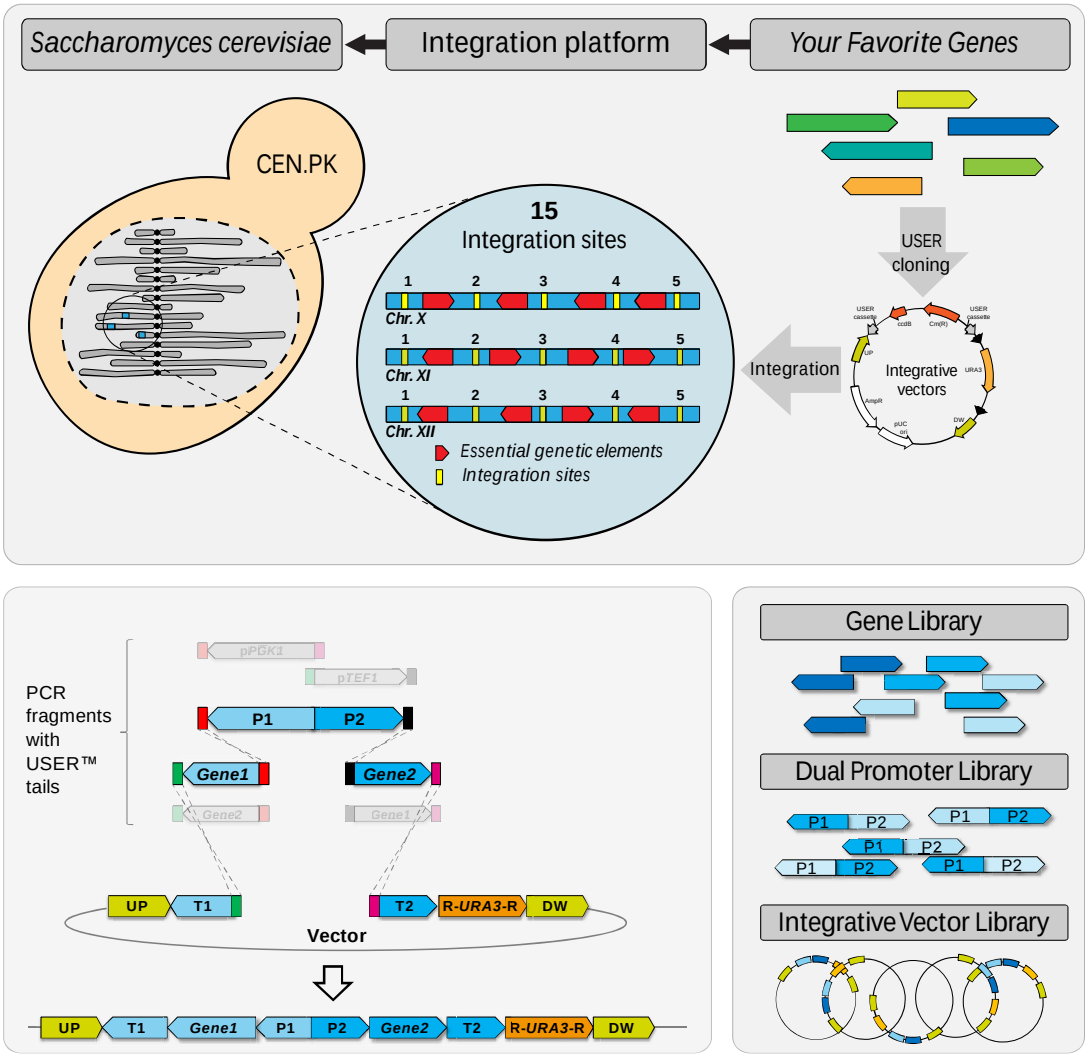


Figure 2.5. Overview of the gene integration platform for yeast *Saccharomyces cerevisiae*. P1, P2 – promoters of interest, T1, T2 – terminators of *ADH1* and *CYC1*, respectively. UP (upstream) and DW (downstream) homology sequences targeting the specific integration sites.

References

[1] J. D. Keasling, “Manufacturing molecules through metabolic engineering.,” *Science*, vol. 330, no. 6009, pp. 1355–8, Dec. 2010.

[2] G. Stephanopoulos and J. Vallino, “Network rigidity and metabolic engineering in metabolite overproduction.,” *Science(Washington)*, vol. 252, pp. 1675–1681, 1991.

[3] J. E. Bailey, “Toward a science of metabolic engineering,” *Science (80-.).*, vol. 252, no. 5013, pp. 1668–1675, 1991.

[4] S. a Benner and a M. Sismour, “Synthetic biology.,” *Nat. Rev. Genet.*, vol. 6, no. 7, pp. 533–43, Jul. 2005.

[5] A. Krivoruchko, V. Siewers, and J. Nielsen, “Opportunities for yeast metabolic engineering: Lessons from synthetic biology.,” *Biotechnol. J.*, vol. 6, no. 3, pp. 262–76, Mar. 2011.

[6] E. Nevoigt, “Progress in metabolic engineering of Saccharomyces cerevisiae.,” *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 3, pp. 379–412, Sep. 2008.

[7] I.-K. Kim, A. Roldão, V. Siewers, and J. Nielsen, “A systems-level approach for metabolic engineering of yeast cell factories.,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 228–48, Mar. 2012.

[8] K.-K. Hong and J. Nielsen, “Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries.,” *Cell. Mol. Life Sci.*, vol. 69, no. 16, pp. 2671–90, Aug. 2012.

[9] A. R. Brochado, C. Matos, B. L. Møller, J. Hansen, U. H. Mortensen, and K. R. Patil, “Improved vanillin production in baker’s yeast through in silico design.,” *Microb. Cell Fact.*, vol. 9, no. 1, p. 84, Jan. 2010.

[10] S. B. Primrose and R. Twyman, *Principles of gene manipulation and genomics*, 7th ed. Oxford: Blackwel, 2006, p. 9.

[11] M. Papini, M. Salazar, and J. Nielsen, “Systems biology of industrial microorganisms,” *Biosyst. Eng. I*, 2010.

[12] K. a Curran and H. S. Alper, “Expanding the chemical palate of cells by combining systems biology and metabolic engineering.,” *Metab. Eng.*, vol. 14, no. 4, pp. 289–97, Jul. 2012.

[13] R. Hatti-Kaul, U. Törnvall, L. Gustafsson, and P. Börjesson, “Industrial biotechnology for the production of bio-based chemicals--a cradle-to-grave perspective.,” *Trends Biotechnol.*, vol. 25, no. 3, pp. 119–24, Mar. 2007.

[14] B. de Jong, V. Siewers, and J. Nielsen, “Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels.,” *Curr. Opin. Biotechnol.*, vol. 23, no. 4, pp. 624–30, Aug. 2012.

[15] J. Nielsen, C. Larsson, A. van Maris, and J. Pronk, “Metabolic engineering of yeast for production of fuels and chemicals.,” *Curr. Opin. Biotechnol.*, vol. 24, no. 3, pp. 398–404, Jun. 2013.

[16] G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. a James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O’Kelly, A. van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. a Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R.

Durbin, and E. J. Louis, “Population genomics of domestic and wild yeasts.,” *Nature*, vol. 458, no. 7236, pp. 337–41, Mar. 2009.

[17] J. a. Barnett, “Beginnings of microbiology and biochemistry: the contribution of yeast research,” *Microbiology*, vol. 149, no. 3, pp. 557–567, Mar. 2003.

[18] H. Feldmann, *Yeast: molecular and cell biology*. Wiley-Blackwell, 2010.

[19] R. K. Mortimer, “Evolution and Variation of the Yeast (Saccharomyces) Genome,” *Genome Res.*, vol. 10, no. 4, pp. 403–409, Apr. 2000.

[20] J. Beggs, “Transformation of yeast by a replicating hybrid plasmid,” *Nature*, vol. 275, pp. 104–109, 1978.

[21] a Hinnen, J. B. Hicks, and G. R. Fink, “Transformation of yeast.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 75, no. 4, pp. 1929–33, Apr. 1978.

[22] A. Goffeau, B. Barrell, H. Bussey, and R. Davis, “Life with 6000 genes,” *Science (80-.).*, vol. 274, no. 5287, 1996.

[23] H. W. Mewes, K. Albermann, M. Bähr, D. Frishman, a Gleissner, J. Hani, K. Heumann, K. Kleine, a Maierl, S. G. Oliver, F. Pfeiffer, and a Zollner, “Overview of the yeast genome.,” *Nature*, vol. 387, no. 6632 Suppl, pp. 7–65, May 1997.

[24] J. L. DeRisi, “Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale,” *Science (80-.).*, vol. 278, no. 5338, pp. 680–686, Oct. 1997.

[25] P. Daran-Lapujade, J.-M. Daran, A. J. A. van Maris, J. H. de Winde, and J. T. Pronk, “Chemostat-based micro-array analysis in baker’s yeast,” *Adv. Microb. Physiol.*, vol. 54, pp. 257–417, 2008.

[26] A. Goffeau, “Four years of post-genomic life with 6,000 yeast genes.,” *FEBS Lett.*, vol. 480, no. 1, pp. 37–41, Aug. 2000.

[27] G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Wang, T. R. Ward, J. Wilhelmy, E. a Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston, “Functional profiling of the Saccharomyces cerevisiae genome.,” *Nature*, vol. 418, no. 6896, pp. 387–91, Jul. 2002.

[28] W. Huh, J. Falvo, L. Gerke, and A. Carroll, “Global analysis of protein localization in budding yeast,” *Nature*, vol. 425, no. 6959, pp. 686–91, Oct. 2003.

[29] A. Baryshnikova, M. Costanzo, Y. Kim, H. Ding, J. Koh, K. Toufighi, J.-Y. Youn, J. Ou, B.-J. San Luis, S. Bandyopadhyay, M. Hibbs, D. Hess, A.-C. Gingras, G. D. Bader, O. G. Troyanskaya, G. W. Brown, B. Andrews, C. Boone, and C. L. Myers, “Quantitative analysis of fitness and genetic interactions in yeast on a genome scale.,” *Nat. Methods*, vol. 7, no. 12, pp. 1017–24, Dec. 2010.

[30] M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Y. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.-Y. Lin, W. Liang, M. Marback, J. Paw, B.-J.

San Luis, E. Shuteriqi, A. H. Y. Tong, N. van Dyk, I. M. Wallace, J. a Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. a Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pál, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A.-C. Gingras, Q. D. Morris, P. M. Kim, C. a Kaiser, C. L. Myers, B. J. Andrews, and C. Boone, "The genetic landscape of a cell.," *Science*, vol. 327, no. 5964, pp. 425–31, Jan. 2010.

[31] S. R. Engel, R. Balakrishnan, G. Binkley, K. R. Christie, M. C. Costanzo, S. S. Dwight, D. G. Fisk, J. E. Hirschman, B. C. Hitz, E. L. Hong, C. J. Krieger, M. S. Livstone, S. R. Miyasato, R. Nash, R. Oughtred, J. Park, M. S. Skrzypek, S. Weng, E. D. Wong, K. Dolinski, D. Botstein, and J. M. Cherry, "Saccharomyces Genome Database provides mutant phenotype data.," *Nucleic Acids Res.*, vol. 38, no. Database issue, pp. D433–6, Jan. 2010.

[32] K. Bitterman, "Longevity regulation in Saccharomyces cerevisiae: linking metabolism, genome stability, and heterochromatin," *… Mol. Biol. …*, vol. 67, no. 3, 2003.

[33] M. Müller, a Heuck, and D. Niessing, "Directional mRNA transport in eukaryotes: lessons from yeast.," *Cell. Mol. Life Sci.*, vol. 64, no. 2, pp. 171–80, Jan. 2007.

[34] L. L. Breeden, "Periodic transcription: a cycle within a cycle.," *Curr. Biol.*, vol. 13, no. 1, pp. R31–8, Jan. 2003.

[35] K. W. Y. Yuen, B. Montpetit, and P. Hieter, "The kinetochore and cancer: what's the connection?," *Curr. Opin. Cell Biol.*, vol. 17, no. 6, pp. 576–82, Dec. 2005.

[36] A. Spradling, B. Ganetsky, and P. Hieter, "New roles for model genetic organisms in understanding and treating human disease: Report From The 2006 Genetics Society of America Meeting," *Genetics*, vol. 172, pp. 2025–2032, 2006.

[37] D. Petranovic and J. Nielsen, "Can yeast systems biology contribute to the understanding of human disease?," *Trends Biotechnol.*, vol. 26, no. 11, pp. 584–90, Nov. 2008.

[38] I. N. Roberts and S. G. Oliver, "The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology.," *Biotechnol. Lett.*, vol. 33, no. 3, pp. 477–87, Mar. 2011.

[39] S. R. Engel and J. M. Cherry, "The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database.," *Database (Oxford).*, vol. 2013, Jan. 2013.

[40] R. K. Mortimer and J. R. Johnston, "Genealogy of principal strains of the yeast genetic stock center.," *Genetics*, vol. 113, no. 1, pp. 35–43, May 1986.

[41] K. Entian and P. Kötter, "Yeast genetic strain and plasmid collections," *Methods Microbiol.*, vol. 36, no. 06, pp. 629–666, 2007.

[42] K. M. Madsen, G. D. B. R. K. Udatha, S. Semba, J. M. Otero, P. Koetter, Y. Ebizuka, T. Kushiro, and G. Panagiotou, "Linking Genotype and Phenotype of Saccharomyces cerevisiae Strains Reveals Metabolic Engineering Targets and Leads to Triterpene Hyper-Producers," vol. 6, no. 3, 2011.

[43] A. B. Canelas, N. Harrison, A. Fazio, J. Zhang, J.-P. Pitkänen, J. van den Brink, B. M. Bakker, L. Bogner, J. Bouwman, J. I. Castrillo, A. Cankorur, P. Chumnanpuen, P. Daran-Lapujade, D. Dikicioglu, K. van Eunen, J. C. Ewald, J. J. Heijnen, B. Kirdar, I. Mattila, F. I. C. Mensonides, A. Niebel, M. Penttilä, J. T. Pronk, M. Reuss, L. Salusjärvi, U. Sauer, D. Sherman, M. Siemann-Herzberg, H. Westerhoff, J. de Winde, D. Petranovic, S. G. Oliver, C. T. Workman, N. Zamboni, and J. Nielsen, "Integrated

multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains.," *Nat. Commun.*, vol. 1, no. 9, p. 145, Jan. 2010.

[44] J. F. Nijkamp, M. van den Broek, E. Datema, S. de Kok, L. Bosman, M. a Luttik, P. Daran-Lapujade, W. Vongsangnak, J. Nielsen, W. H. M. Heijne, P. Klaassen, C. J. Paddon, D. Platt, P. Kötter, R. C. van Ham, M. J. T. Reinders, J. T. Pronk, D. de Ridder, and J.-M. Daran, "*De novo* sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology.," *Microb. Cell Fact.*, vol. 11, no. 1, p. 36, Jan. 2012.

[45] J. M. Otero and J. Nielsen, "Industrial systems biology.," *Biotechnol. Bioeng.*, vol. 105, no. 3, pp. 439–60, Feb. 2010.

[46] M. S. Siddiqui, K. Thodey, I. Trenchard, and C. D. Smolke, "Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools.," *FEMS Yeast Res.*, vol. 12, no. 2, pp. 144–70, Mar. 2012.

[47] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. a Ho, R. a Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling, "Production of the antimalarial drug precursor artemisinic acid in engineered yeast.," *Nature*, vol. 440, no. 7086, pp. 940–3, Apr. 2006.

[48] P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman, and C. J. Paddon, "Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 3, pp. E111–8, Jan. 2012.

[49] C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, a Tai, a Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, and J. D. Newman, "High-level semi-synthetic production of the potent antimalarial artemisinin.," *Nature*, vol. 496, no. 7446, pp. 528–32, Apr. 2013.

[50] S. Watanabe, A. Abu Saleh, S. P. Pack, N. Annaluru, T. Kodaki, and K. Makino, "Ethanol production from xylose by recombinant Saccharomyces cerevisiae expressing protein-engineered NADH-preferring xylose reductase from Pichia stipitis.," *Microbiology*, vol. 153, no. Pt 9, pp. 3044–54, Sep. 2007.

[51] V. Guadalupe Medina, M. J. H. Almering, A. J. a van Maris, and J. T. Pronk, "Elimination of glycerol production in anaerobic cultures of a Saccharomyces cerevisiae strain engineered to use acetic acid as an electron acceptor.," *Appl. Environ. Microbiol.*, vol. 76, no. 1, pp. 190–5, Jan. 2010.

[52] K. O. Yu, S. W. Kim, and S. O. Han, "Engineering of glycerol utilization pathway for ethanol production by Saccharomyces cerevisiae.," *Bioresour. Technol.*, vol. 101, no. 11, pp. 4157–61, Jun. 2010.

[53] E. J. Steen, R. Chan, N. Prasad, S. Myers, C. J. Petzold, A. Redding, M. Ouellet, and J. D. Keasling, "Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol.," *Microb. Cell Fact.*, vol. 7, p. 36, Jan. 2008.

[54] X. Chen, K. F. Nielsen, I. Borodina, M. C. Kielland-Brandt, and K. Karhumaa, "Increased isobutanol production in Saccharomyces cerevisiae by overexpression of genes in valine metabolism," *Biotechnol Biofuels*, vol. 4, no. 21, pp. 2089–2090, 2011.

[55] P. P. Peralta-Yahya, M. Ouellet, R. Chan, A. Mukhopadhyay, J. D. Keasling, and T. S. Lee, “Identification and microbial production of a terpene-based advanced biofuel,” *Nat. Commun.*, vol. 2, p. 483, 2011.

[56] K. O. Yu, J. Jung, S. W. Kim, C. H. Park, and S. O. Han, “Synthesis of FAEEs from glycerol in engineered Saccharomyces cerevisiae using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase,” *Biotechnol. Bioeng.*, vol. 109, no. 1, pp. 110–115, 2012.

[57] J.-F. Liu, K.-L. Nie, L.-H. Fan, F. Wang, T.-W. Tan, and L. Deng, “Increased production of FAEEs for biodiesel with lipase enhanced Saccharomyces cerevisiae,” *Process Biochem.*, vol. 48, no. 8, pp. 1212–1215, Aug. 2013.

[58] K. M. Overkamp, B. M. Bakker, P. Kötter, M. A. H. Luttik, J. P. Van Dijken, T. Jack, and J. T. Pronk, “Metabolic Engineering of Glycerol Production in Saccharomyces cerevisiae Metabolic Engineering of Glycerol Production in Saccharomyces cerevisiae,” 2002.

[59] A. J. A. van Maris, J.-M. A. Geertman, A. Vermeulen, M. K. Groothuizen, A. A. Winkler, M. D. W. Piper, J. P. van Dijken, and J. T. Pronk, “Directed evolution of pyruvate decarboxylase-negative Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast,” *Appl. Environ. Microbiol.*, vol. 70, no. 1, pp. 159–166, 2004.

[60] B. Zhang, R. Carlson, and F. Srienc, “Engineering the Monomer Composition of Polyhydroxyalkanoates Synthesized in Saccharomyces cerevisiae Engineering the Monomer Composition of Polyhydroxyalkanoates Synthesized in Saccharomyces cerevisiae,” vol. 72, no. 1, 2006.

[61] M. H. Toivari, H. Maaheimo, M. Penttilä, and L. Ruohonen, “Enhancing the flux of D-glucose to the pentose phosphate pathway in Saccharomyces cerevisiae for the production of D-ribose and ribitol,” *Appl. Microbiol. Biotechnol.*, vol. 85, no. 3, pp. 731–739, 2010.

[62] A. M. Raab, G. Gebhardt, N. Bolotina, D. Weuster-Botz, and C. Lang, “Metabolic engineering of Saccharomyces cerevisiae for the biotechnological production of succinic acid.,” *Metab. Eng.*, vol. 12, no. 6, pp. 518–25, Nov. 2010.

[63] L. Zhao, J. Wang, J. Zhou, L. Liu, G. Du, and J. Chen, “[Modification of carbon flux in Sacchromyces cerevisiae to improve L-lactic acid production].,” *Wei sheng wu xue bao= Acta Microbiol. Sin.*, vol. 51, no. 1, pp. 50–58, 2011.

[64] K. Kocharin, V. Siewers, and J. Nielsen, “Improved polyhydroxybutyrate production by Saccharomyces cerevisiae through the use of the phosphoketolase pathway.,” *Biotechnol. Bioeng.*, vol. 110, no. 8, pp. 2216–24, Aug. 2013.

[65] J. Becker, G. Armstrong, M. Vandermerwe, M. Lambrechts, M. Vivier, and I. Pretorius, “Metabolic engineering of for the synthesis of the wine-related antioxidant resveratrol,” *FEMS Yeast Res.*, vol. 4, no. 1, pp. 79–85, Oct. 2003.

[66] J. R. Millis, J. Maurina-Brunker, and T. W. McMullin, “Production of farnesol and geranylgeraniol.” Google Patents, 2004.

[67] M. Sauer, P. Branduardi, M. Valli, and D. Porro, “Production of l-Ascorbic Acid by Metabolically Engineered Saccharomyces cerevisiae and Zygosaccharomyces bailii Production of L -Ascorbic Acid by Metabolically Engineered Saccharomyces cerevisiae and Zygosaccharomyces bailii,” 2004.

[68] R. Verwaal, J. Wang, J.-P. Meijnen, H. Visser, G. Sandmann, J. a van den Berg, and A. J. J. van Ooyen, “High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous.,” *Appl. Environ. Microbiol.*, vol. 73, no. 13, pp. 4342–50, Jul. 2007.

[69] E. E. H. Hansen, B. B. L. Møller, G. R. G. Kock, C. M. Bünner, C. Kristensen, O. R. Jensen, F. T. Okkels, C. E. Olsen, M. S. Motawia, and J. Hansen, "*De novo* biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (Saccharomyces cerevisiae).," *Appl. Environ. Microbiol.*, vol. 75, no. 9, pp. 2765–74, May 2009.

[70] V. Siewers, R. San-Bento, and J. Nielsen, "Implementation of communication-mediating domains for non-ribosomal peptide production in Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 106, no. 5, pp. 841–4, Aug. 2010.

[71] J. Kirby, M. Nishimoto, J. G. Park, S. T. Withers, F. Nowroozi, D. Behrendt, E. J. G. Rutledge, J. L. Fortman, H. E. Johnson, J. V Anderson, and J. D. Keasling, "Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in Saccharomyces cerevisiae.," *Phytochemistry*, vol. 71, no. 13, pp. 1466–73, Sep. 2010.

[72] L. Albertsen, Y. Chen, L. S. Bach, S. Rattleff, J. Maury, S. Brix, J. Nielsen, and U. H. Mortensen, "Diversion of flux toward sesquiterpene production in Saccharomyces cerevisiae by fusion of host and heterologous enzymes.," *Appl. Environ. Microbiol.*, vol. 77, no. 3, pp. 1033–40, Feb. 2011.

[73] S. Tavares, T. Grotkjær, T. Obsen, R. P. Haslam, J. a Napier, and N. Gunnarsson, "Metabolic engineering of Saccharomyces cerevisiae for production of Eicosapentaenoic Acid, using a novel {Delta}5-Desaturase from Paramecium tetraurelia.," *Appl. Environ. Microbiol.*, vol. 77, no. 5, pp. 1854–61, Mar. 2011.

[74] F. Koopman, J. Beekwilder, B. Crimi, A. van Houwelingen, R. D. Hall, D. Bosch, A. J. a van Maris, J. T. Pronk, and J.-M. Daran, "*De novo* production of the flavonoid naringenin in engineered Saccharomyces cerevisiae.," *Microb. Cell Fact.*, vol. 11, p. 155, Jan. 2012.

[75] M. a Asadollahi, J. Maury, K. R. Patil, M. Schalk, A. Clark, and J. Nielsen, "Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering.," *Metab. Eng.*, vol. 11, no. 6, pp. 328–34, Nov. 2009.

[76] M. a Asadollahi, J. Maury, M. Schalk, A. Clark, and J. Nielsen, "Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 106, no. 1, pp. 86–96, May 2010.

[77] A. R. Brochado and K. R. Patil, "Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering.," *Biotechnol. Bioeng.*, vol. 110, no. 2, pp. 656–9, Feb. 2013.

[78] J. Nielsen and M. C. Jewett, "Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae.," *FEMS Yeast Res.*, vol. 8, no. 1, pp. 122–31, Feb. 2008.

[79] T. Osterlund, I. Nookaew, and J. Nielsen, "Fifteen years of large scale metabolic modeling of yeast: developments and impacts.," *Biotechnol. Adv.*, vol. 30, no. 5, pp. 979–88, 2012.

[80] N. a Da Silva and S. Srikrishnan, "Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae.," *FEMS Yeast Res.*, vol. 12, no. 2, pp. 197–214, Mar. 2012.

[81] S. Partow, V. Siewers, S. Bjørn, J. Nielsen, and J. Maury, "Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae," *Yeast*, no. July, pp. 955–964, 2010.

[82] J. Sun, Z. Shao, H. Zhao, N. Nair, F. Wen, J.-H. Xu, and H. Zhao, "Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 109, no. 8, pp. 2082–92, Aug. 2012.

[83] H. Alper and C. Fischer, “Tuning genetic control through promoter engineering,” *Proc. …*, vol. 103, no. 36, 2005.

[84] B. a Blount, T. Weenink, S. Vasylechko, and T. Ellis, “Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology.,” *PLoS One*, vol. 7, no. 3, p. e33279, Jan. 2012.

[85] F. S. Hartner, C. Ruth, D. Langenegger, S. N. Johnson, P. Hyka, G. P. Lin-Cereghino, J. Lin-Cereghino, K. Kovar, J. M. Cregg, and A. Glieder, “Promoter library designed for fine-tuned gene expression in Pichia pastoris.,” *Nucleic Acids Res.*, vol. 36, no. 12, p. e76, Jul. 2008.

[86] T. S. Bayer and C. D. Smolke, “Programmable ligand-controlled riboregulators of eukaryotic gene expression.,” *Nat. Biotechnol.*, vol. 23, no. 3, pp. 337–43, Mar. 2005.

[87] M. H. Medema, R. van Raaphorst, E. Takano, and R. Breitling, “Computational tools for the synthetic design of biochemical pathways.,” *Nat. Rev. Microbiol.*, vol. 10, no. 3, pp. 191–202, Mar. 2012.

[88] M. D. Mikkelsen, L. D. Buron, B. Salomonsen, C. E. Olsen, B. G. Hansen, U. H. Mortensen, and B. A. Halkier, “Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform.,” *Metab. Eng.*, vol. 14, no. 2, pp. 104–11, Mar. 2012.

[89] K. L. J. Prather and C. H. Martin, “*De novo* biosynthetic pathways: rational design of microbial chemical factories.,” *Curr. Opin. Biotechnol.*, vol. 19, no. 5, pp. 468–74, Oct. 2008.

[90] L. Clarke and J. Carbon, “Isolation of a yeast centromere and construction of functional small circular chromosomes,” *Nature*, vol. 287, pp. 504–509, 1980.

[91] a B. Futcher and B. S. Cox, “Maintenance of the 2 microns circle plasmid in populations of Saccharomyces cerevisiae.,” *J. Bacteriol.*, vol. 154, no. 2, pp. 612–22, May 1983.

[92] M. a Romanos, C. a Scorer, and J. J. Clare, “Foreign gene expression in yeast: a review.,” *Yeast*, vol. 8, no. 6, pp. 423–88, Jun. 1992.

[93] H. Ma, S. Kunes, P. J. Schatz, and D. Botstein, “Plasmid construction by homologous recombination in yeast,” *Gene*, vol. 58, no. 2, pp. 201–216, 1987.

[94] R. D. Gietz and S. Akio, “New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites,” *Gene*, vol. 74, no. 2, pp. 527–534, Dec. 1988.

[95] R. S. Sikorski and P. Hieter, “A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in,” *Genetics*, no. 122, pp. 19–27, 1989.

[96] T. W. Christianson, R. S. Sikorski, M. Dante, J. H. Shero, and P. Hieter, “Multifunctional yeast high-copy-number shuttle vectors,” *Gene*, vol. 110, no. 1, pp. 119–122, 1992.

[97] C. B. Brachmann, a Davies, G. J. Cost, E. Caputo, J. Li, P. Hieter, and J. D. Boeke, “Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications.,” *Yeast*, vol. 14, no. 2, pp. 115–32, Jan. 1998.

[98] C. Taxis and M. Knop, “System of centromeric, episomal, and integrative vectors based on drug resistance markers for Saccharomyces cerevisiae,” *Biotechniques*, vol. 40, no. 1, pp. 73–78, Jan. 2006.

[99] D. Mumberg, R. Müller, and M. Funk, "Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression.," *Nucleic Acids Res.*, vol. 22, no. 25, pp. 5767–5768, 1994.

[100] D. Mumberg, R. Müller, and M. Funk, "Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.," *Gene*, vol. 156, no. 1, pp. 119–22, Apr. 1995.

[101] E. Garí, L. Piedrafita, M. Aldea, and E. Herrero, "A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae.," *Yeast*, vol. 13, no. 9, pp. 837–48, Jul. 1997.

[102] C. Miller, M. Martinat, and L. Hyman, "Assessment of aryl hydrocarbon receptor complex interactions using pBEVY plasmids: expression vectors with bi-directional promoters for use in Saccharomyces," *Nucleic Acids Res.*, vol. 26, no. 15, pp. 3577–3583, 1998.

[103] J. W. Lee, H. U. Kim, S. Choi, J. Yi, and S. Y. Lee, "Microbial production of building block chemicals and polymers.," *Curr. Opin. Biotechnol.*, vol. 22, no. 6, pp. 758–67, Dec. 2011.

[104] J. Maury, M. a Asadollahi, K. Møller, M. Schalk, A. Clark, L. R. Formenti, and J. Nielsen, "Reconstruction of a bacterial isoprenoid biosynthetic pathway in Saccharomyces cerevisiae.," *FEBS Lett.*, vol. 582, no. 29, pp. 4032–8, Dec. 2008.

[105] I. Sadowski, T. Su, and J. Parent, "Disintegrator vectors for single-copy yeast chromosomal integration," *Yeast*, no. February, pp. 447–455, 2007.

[106] S. Alberti, A. A. D. Gitler, and S. Lindquist, "A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae," *Yeast*, no. June, pp. 913–919, 2007.

[107] A. Li, Z. Liu, Q. Li, L. Yu, D. Wang, and X. Deng, "Construction and characterization of bidirectional expression vectors in Saccharomyces cerevisiae.," *FEMS Yeast Res.*, vol. 8, no. 1, pp. 6–9, Feb. 2008.

[108] F. Fang, K. Salmon, M. W. Y. Shen, K. A. Aeling, E. Ito, B. Irwin, U. P. C. Tran, G. W. Hatfield, N. A. Da Silva, and S. Sandmeyer, "A vector set for systematic metabolic engineering in Saccharomyces cerevisiae," *Yeast*, vol. 3, no. October 2010, pp. 123–136, 2011.

[109] M. Shen, F. Fang, S. Sandmeyer, and N. Da Silva, "Development and characterization of a vector set with regulated promoters for systematic metabolic engineering in Saccharomyces cerevisiae," *Yeast*, no. November, pp. 495–503, 2012.

[110] N. B. Jensen, T. Strucko, K. R. Kildegaard, F. David, J. Maury, U. H. Mortensen, J. Förster, J. B. Nielsen, and I. Borodina, "EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae," *FEMS Yeast Res.*, pp. DOI:10.1111/1567–1364.12118, 2013.

[111] D. J. Mead, D. C. J. Gardner, and S. G. Oliver, "The yeast 2 μ plasmid: strategies for the survival of a selfish DNA," *Mol. Gen. Genet. MGG*, vol. 205, no. 3, pp. 417–421, 1986.

[112] B. Futcher and J. Carbon, "Toxic effects of excess cloned centromeres.," *Mol. Cell. Biol.*, vol. 6, no. 6, 1986.

[113] A. S. Karim, K. a Curran, and H. S. Alper, "Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications.," *FEMS Yeast Res.*, vol. 13, no. 1, pp. 107–16, Feb. 2013.

[114] T. Orr-Weaver and J. Szostak, "Fungal recombination.," *Microbiol. Rev.*, vol. 49, no. 1, pp. 33–58, 1985.

[115] U. Gueldener, J. Heinisch, G. J. Koehler, D. Voss, and J. H. Hegemann, “A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.,” *Nucleic Acids Res.*, vol. 30, no. 6, p. e23, Mar. 2002.

[116] H. Dhamankar and K. L. J. Prather, “Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals.,” *Curr. Opin. Struct. Biol.*, vol. 21, no. 4, pp. 488–94, Aug. 2011.

[117] D. B. Flagfeldt, V. Siewers, L. Huang, and J. Nielsen, “Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae,” *Yeast*, no. August, pp. 545–551, 2009.

[118] M. C. Lorenz, R. S. Muir, E. Lim, J. McElver, S. C. Weber, and J. Heitman, “Gene disruption with PCR products in Saccharomyces cerevisiae,” *Gene*, vol. 158, no. 1, pp. 113–117, Jan. 1995.

[119] D. C. Amberg, D. Botstein, and E. M. Beasley, “Precise gene disruption in Saccharomyces cerevisiae by double fusion polymerase chain reaction.,” *Yeast*, vol. 11, no. 13, pp. 1275–80, Oct. 1995.

[120] C. Raymond, T. Pownder, and S. Sexson, “General method for plasmid construction using homologous recombination,” *Biotechniques*, vol. 141, no. January, 1999.

[121] K. R. Oldenburg, K. T. Vo, S. Michaelis, and C. Paddon, “Recombination-mediated PCR-directed plasmid construction in vivo in yeast.,” *Nucleic Acids Res.*, vol. 25, no. 2, pp. 451–2, Jan. 1997.

[122] Z. Shao, H. Zhao, and H. Zhao, “DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways.,” *Nucleic Acids Res.*, vol. 37, no. 2, p. e16, Feb. 2009.

[123] J. L. Hartley, G. F. Temple, and M. A. Brasch, “DNA Cloning Using In Vitro Site-Specific Recombination,” *Genome Res.*, vol. 10, no. 11, pp. 1788–1795, Nov. 2000.

[124] D. D. G. Gibson, L. Young, R. R.-Y. Chuang, J. C. Venter, C. A. 3rd Hutchison, and H. O. Smith, “Enzymatic assembly of DNA molecules up to several hundred kilobases,” *Nat. Methods*, vol. 6, no. 5, pp. 343–345, 2009.

[125] C. Engler, R. Kandzia, and S. Marillonnet, “A one pot, one step, precision cloning method with high throughput capability.,” *PLoS One*, vol. 3, no. 11, p. e3647, Jan. 2008.

[126] T. Wang, X. Ma, H. Zhu, A. Li, G. Du, and J. Chen, “Available methods for assembling expression cassettes for synthetic biology.,” *Appl. Microbiol. Biotechnol.*, vol. 93, no. 5, pp. 1853–63, Mar. 2012.

[127] H. H. Nour-Eldin, B. G. Hansen, M. H. H. Nørholm, J. K. Jensen, and B. a Halkier, “Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments.,” *Nucleic Acids Res.*, vol. 34, no. 18, p. e122, Jan. 2006.

[128] F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, and B. a Halkier, “USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.,” *Nucleic Acids Res.*, vol. 35, no. 7, p. e55, Jan. 2007.

[129] J. M. Walker, H. H. Nour-Eldin, F. Geu-Flores, and B. A. Halkier, “USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories,” in in *Plant Secondary Metabolism Engineering*, Springer, 2010, pp. 185–200.

[130] L. M. Wingler and V. W. Cornish, “Reiterative Recombination for the in vivo assembly of libraries of multigene pathways.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 37, pp. 15135–40, Sep. 2011.

[131] N. G. a Kuijpers, S. Chroumpi, T. Vos, D. Solis-Escalante, L. Bosman, J. T. Pronk, J.-M. Daran, and P. Daran-Lapujade, “One step assembly and targeted integration of multi-gene constructs assisted by the I-SceI meganuclease in Saccharomyces cerevisiae.,” *FEMS Yeast Res.*, Sep. 2013.

[132] B. G. Hansen, B. Salomonsen, M. T. Nielsen, J. B. Nielsen, N. B. Hansen, K. F. Nielsen, T. B. Regueira, J. Nielsen, K. R. Patil, and U. H. Mortensen, “Versatile enzyme expression and characterization system for Aspergillus nidulans, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case.,” *Appl. Environ. Microbiol.*, vol. 77, no. 9, pp. 3044–51, May 2011.

Chapter 3

Rapid pathways assembly in yeast

Title[1]

EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*

Authors

N. B. Jensen[1,2,] **T. Strucko**[2], K. R. Kildegaard[1], F. David[3], J. Maury[1], U. H. Mortensen[2], J. Förster[1], J. B. Nielsen[1,3] and I. Borodina[1]

Affiliations

1-The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle Allé 6, Hørsholm DK-2970, Denmark

2-Department of Systems Biology, Technical University of Denmark, Søltofts Plads Building 223, Kgs Lyngby DK-2800, Denmark

3-Department of Chemical and Biological Engineering, Chalmers University of Technology, SE412 96 Gothenburg, Sweden

Keywords: Genome editing, *Saccharomyces cerevisiae*, metabolic engineering, integrative vectors, USER cloning

[1] Published in FEMS Yeast Research DOI: 10.1111/1567-1364.12118

[2] Present address: Evolva Biotech A/S, Lersø Parkallé 42-44, DK-2100 København Ø, Denmark

Abstract

Development of strains for efficient production of chemicals and pharmaceuticals requires multiple rounds of genetic engineering. In this study we describe construction and characterization of EasyClone vector set for baker's yeast *Saccharomyces cerevisiae*, which enables simultaneous expression of multiple genes with an option of recycling selection markers. The vectors combine the advantage of efficient uracil-excision reaction based cloning and Cre-Lox mediated marker recycling system. The episomal and integrative vector sets were tested by inserting genes encoding cyan, yellow and red fluorescent proteins into separate vectors and analyzing for co-expression of proteins by flow cytometry. Cells expressing genes encoding for the three fluorescent proteins from three integrations exhibited a much higher level of simultaneous expression than cells producing fluorescent proteins encoded on episomal plasmids, where correspondingly 95% and 6% of the cells were within a fluorescence interval of Log_{10} mean ± 15% for all three colors. We demonstrate that selective markers can be simultaneously removed using Cre-mediated recombination and all the integrated heterologous genes remain in the chromosome and show unchanged expression levels. Hence this system is suitable for metabolic engineering in yeast where multiple rounds of gene introduction and marker recycling can be carried out.

Introduction

Baker's yeast *S. cerevisiae* is an attractive cell factory for industrial biotechnology [1]. It is used for the production of food and beverages and for chemicals, enzymes and pharmaceuticals. Due to extensive efforts within yeast genetic research, a vast number of genetic and molecular tools have been developed (for reviews see [2]–[4]. Among these tools high- and low- copy as well as integrative plasmids have found extensive use in gene function studies and in metabolic engineering [5]. There are strengths and weaknesses for each type of plasmids and eventually the choice depends on the overall goal. When it comes to metabolic engineering of yeast to obtain a process with high titer, rate and yield, multiple rounds of strain engineering are commonly required. To cut down the costs it is important that the turnaround time of the metabolic engineering cycle is as short as possible. At the same time it is critical that the cycle is at the highest possible standard e.g. in terms of stability of expression of the genes introduced and that expression levels can be controlled in a reliable way either by inducible promoters or constitutive promoters of various strengths (see [2] for references and details). Another important consideration for pathway engineering is the ability to co-express the introduced genes and at the desired levels in each and every cell of the yeast population. This is a problem as the copy number for both the high- and low copy-number plasmids fluctuates in the cell population [6], [7]. The stability issue can be overcome by using integration plasmids, where the expression cassettes are integrated in the genome. Several integration vector series have been developed over time [8]–[11]. Despite the stable nature of chromosomal integrations when compared to e.g. high copy episomal plasmids, instability can occur if the introduced fragments share a high degree of sequence homology or if insertions are multiple tandem insertions. If the latter is the case, there is a high risk of chromosomal rearrangements including loss of the introduced genes due to direct repeat recombination [12], [13].

Another crucial step for the turnaround time of a metabolic engineering cycle is the cloning phase. Several high-throughput cloning methods developed over time have proven to be of great importance, e.g. Gateway™ cloning (Invitrogen) [14], Gibson Assembly® cloning (New England Biolabs) [15], Golden Gate cloning [16] and Infusion cloning from Clontech. Another method is the uracil-specific excision reaction (USER)-based cloning technique [17].This cloning technique was the basic technique for the plasmid set developed by [18], which allows stable integration into 15 individual integration sites, where each site was validated for growth impairment and production of β-galactosidase. Furthermore, the insertion sites on each chromosome are interspaced by essential genetic elements preventing loop out of the inserted fragments by homologous recombination. As an example of their system's applicability the authors showed successful

expression of a complex eight gene indole glucosinolate biosynthetic pathway in *S. cerevisiae*. One limitation of this plasmid set is the fact that it is based on only one selectable marker, *Kluyveromyces lactis URA3*, which needs to be recycled during sequential integration steps in a process mediated by direct repeat recombination and 5-fluoroorotic acid selection. Hence, introduction of multi-gen pathways will be time consuming as it will require many rounds of strain transformation and marker elimination.

Our intention has been to create a method that allows repeated cycles of genetic engineering, in which multiple genes are simultaneously stably integrated into the genome of *S. cerevisiae*. We describe integrative vector set EasyClone with a wide repertoire of LoxP-flanked selection markers, developed on the basis of [18] vectors. As proof of concept, we simultaneously integrate three different gene targeting cassettes containing genes encoding three different fluorescent proteins and then loop-out the markers without losing fluorescent proteins genes. We also evaluate the heterogeneity in the population of cells expressing multiple proteins from the integrative EasyClone vectors and from 2µ-based episomal plasmids.

Materials and Methods

Strains and media

Saccharomyces cerevisiae CEN.PK102-5B (MAT*a ura3-52 his3Δ1 leu2-3/112 MAL2-8^c SUC2*) strain was obtained from Verena Siewers (Chalmers University). Yeast transformants were selected on synthetic complete (SC) drop out media lacking the amino acids matching the auxotrophic markers on the plasmids used. These SC plates were made from pre-mixed drop-out powders from Sigma-Aldrich. When yeast was grown in liquid media it was either in SC, Delft or standard yeast peptone dextrose (YPD) media. Delft contained per liter: 7.5 g $(NH_4)_2SO_4$, 14.4g KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 22 g dextrose, 2 mL trace metals solution, and 1 mL vitamins. The pH of Delft medium was adjusted to 6 prior to autoclavation. Vitamin solution was added to Delft medium after autoclavation. The trace metals solution contained per liter: 4.5 g $CaCl_2 \cdot 2H_2O$, 4.5 g $ZnSO_4 \cdot 7H_2O$, 3 g $FeSO_4 \cdot 7H_2O$, 1 g H_3BO_3, 1 g $MnCl_2 \cdot 4H_2O$, 0.4 g $Na_2MoO_4 \cdot 2H_2O$, 0.3 g $CoCl_2 \cdot 6H_2O$, 0.1 g $CuSO_4 \cdot 5H_2O$, 0.1 g KI, 15 g EDTA. The trace metals solution was prepared by dissolving all the components except EDTA in 900 mL ultra-pure water at pH 6. The solution was then gently heated and EDTA was added. In the end, the pH was adjusted to 4, and the solution volume was adjusted to 1L and autoclaved (121^{o}C in 20 min). This solution was stored at +4^{o}C. The vitamins solution had per liter: 50 mg biotin, 200 mg p-aminobenzoic acid, 1 g nicotinic acid, 1 g Ca-pantotenate, 1 g pyridoxine-HCl, 1 g thiamine-HCl, 25 g myo-inositol. Biotin was dissolved in 20 mL 0.1 M NaOH and 900 mL water is added. pH was adjusted to 6.5 with HCl and the rest of the vitamins was added. pH was re-adjusted to 6.5 just

before and after adding m-inositol. The final volume was adjusted to 1 L and sterile-filtered before storage at +4°C.

All standard cloning was carried out using *Escherichia coli* strain DH5α, which was grown in standard Lysogeny broth (LB) medium containing 100 μg mL^{-1} of ampicillin. For the cloning of plasmid carrying the *ccdB* gene and chloramphenicol cassette, *E. coli ccdB* strain was used as a host strain and transformants were selected on LB medium containing 100 μg mL^{-1} ampicillin and 25 μg mL^{-1} chloramphenicol.

Plasmid and strain construction

The episomal plasmids were generated as follows: the 1.8-kb fragment carrying the USER cassette, *ccdB* gene and chloramphenicol marker was generated by PCR amplification using primers pESC_U_ccdB_fw and pESC_U_ccdB_rv and plasmid pCfB49 (pXII-1-ccdB) as a template. The PCR fragment was digested with SacI and XhoI, gel-purified and then ligated into plasmid pESC-URA or pESC-HIS, which were digested with the same enzyme pair. The final plasmids were designated as pCfB54 (pESC-URA-ccdB-USER) and pCfB55 (pESC-HIS-ccdB-USER), respectively. Finally, the pCfB54 and pCfB55 plasmids were digested with FastDigest® AsiSI to remove the *ccdB* gene including the chloramphenicol marker and re-ligated to generate the final plasmids pCfB132 (pESC-URA-USER) and pCfB291 (pESC-HIS-USER).

To construct pCfB220 (pESC-LEU-USER), the 36 bp fragment carrying the USER cassette was excised from the plasmid pCfB132 (pESC-URA-USER) using SacI and XhoI, gel-purified and then ligated into plasmid pESC-LEU, which was digested with the same enzyme pair.

The integration plasmids were made from the plasmid set previously described in [18] by replacing the directed repeats (DR) flanking *K. lactis URA3* selection marker with different selection markers flanked with *LoxP* sites [19]. The selection marker exchange was accomplished by uracil-specific excision reaction (USER) [17], where the parent plasmids and the different *LoxP* flanked selection marker fragments were PCR amplified by PfuX7 polymerase [20] using oligos listed in **Table 3.1**. The integration plasmids listed in **Table 3.2**. using the following protocol: 3 μL of gel purified plasmid PCR product was mixed with 5 μL of gel purified selection marker PCR fragment together with 1 μL Taq polymerase buffer, and 1 μL USER enzyme (EB). The mix was incubated at 37°C for 25 min, at 25°C for 25 min and transformed into chemically competent *E. coli* DH5α cells. The clones with correct inserts were identified by colony PCR and the plasmids were isolated from overnight *E. coli* cultures and confirmed by sequencing. This way the following plasmids were obtained: pCfB255, pCfB257, pCfB258, pCfB259, pCfB260, pCfB261, pCfB262, pCfB353, pCfB388, pCfB389, pCfB390, pCfB391.

Plasmids expressing genes of fluorescent protein were constructed using USER-cloning as previously described in [21]. Prior cloning, episomal and integrative vectors containing USER cassettes were digested with AsiSI and subsequently with the nicking endonuclease Nb.BsmI **Figure 3.1**. Each batch of USER vector prepared for USER cloning (**SI Figure S 3.1**) was tested for number of background transformants, i.e. the number of transformants growing on selective plates but that do not carry a vector with the insert. This experiment was carried as follows. A defined amount of prepared USER vector (~30 ng) was mixed with a PCR product bearing the green fluorescent protein encoding gene under the control of an *E. coli* promoter. This PCR product was generated by using primers ID1493 and ID1494 and pCfB774 as template. Prepared USER vector and PCR product were mixed in a 1:3 vector to insert molar ratio. After USER reaction and transformation of chemically competent *E. coli* cells, cells were spread on LB_{AMP} and incubated for approximately 18h at 37^{o}C. The plate was directly analyzed under blue light excitation (bench top blue light table): the number of white colonies corresponds to the number of "background" transformants, while the fluorescent colonies are the "positive" transformants. Routinely at least 80-90% of the colonies were fluorescent.

The coding sequences of the genes encoding for the three different fluorescence proteins and *TEF1* promoter were amplified by PCR using PfuX7 polymerase and primers listed in **Table 3.1**. *CFP*, *RFP* and *YFP* were obtained from appropriate plasmid templates pWJ1163, pWJ1350 and pWJ1165, respectively, and the promoter *TEF1* from genomic DNA of the CEN.PK113-11C strain. The promoter *TEF1* and cDNAs were cloned into previously linearized and tested expression vectors. A total of six constructs were produced: three designed to integrate on chromosome X and three to be expressed episomally. All constructed plasmids were validated by sequencing.

Yeast strain CEN.PK102-5B was transformed with different combinations of either episomal plasmids or linearized fragments for chromosomal integration by the lithium acetate transformation protocol [22]. Prior transformation, integrative plasmids were digested with NotI and the fragment containing the sequences for integration was purified from agarose gel. For each integrative fragment we used 300-700 ng DNA for each transformation. For episomal plasmids 200-400 ng DNA was used per transformation. The integration of the plasmids was verified by PCR analysis on yeast gDNA purified by the ZR Fungal/Bacterial DNA MiniPrep^{TM} kit from Zymo Research using primers listed in **Table 3.1**.

Table 3.1. List of the primers used for vector construction and strain verification. USER-specific overhangs are marked in **bold**, translational enhancer (Kozak) sequence is <u>underlined</u> (Cavener & Ray, 1991; Nakagawa, Niimura, Gojobori, Tanaka, & Miura, 2008)

Name	Sequence (5' to 3')	Application
pESC_U_ccdB_fw	AAAAGAGCTCGAATGCGTGCGATCGCAG	Amplification of USER cassette, ccdB gene and chloramphenicol cassette
pESC_U_ccdB_rv	AAAACTCGAGGAATGCACGCGATCGCTG	
ID399USERrev	**ATTGGGU**GCATAGGCCACTAGTGGATCTG	Amplification of LoxP flanked selection marker cassettes
ID400USERfwd	**ATCGCGU**CAGCTGAAGCTTCGTACGC	
ID401pIntFwdU	**ACCCAAU**TCGCCCTATAGTGAGTCG	Amplification of integrative plasmid backbone
ID402pintRevU	**ACGCGAU**CTTCGAGCGTCCCAAAACC	
ID1493	**CGTGCGAU**CCGCATAGGGAGTGTAAATTTATC	Positive GFP control fragment for USER plasmid
ID1494	**CACGCGAU**AGTGAAAGGAAGGCCCATGAG	
PTEF1_fw	**ACCTGCACU**TTGTAATTAAAACTTAG	Amplification of TEF1 promoter
PTEF1_rv	**CACGCGAU**GCACACACCATAGCTTC	
YFP/CFP_F+	**AGTGCAGGU**<u>AAAACA</u>ATGAGTAAAGGAGAAGAACTTTTCAC	Amplification of YFP and CFP genes
YFP/CFP_R+	**CGTGCGAU**TCATTTGTATAGTTCATCCATGCCATG	
RFP_F+	**AGTGCAGGU**<u>AAAACA</u>ATGGCCTCCTCCGAGGACGTCATC	Amplification of RFP gene
RFP_R+	**CGTGCGAU**TCAGGCGCCGGTGGAGTGGCGG	
ID901 X-2-up-out	TGCGACAGAAGAAAGGGAAG	PCR with ID339 verifies insertion in X-2-UP
ID902-X-2-down-out	GAGAACGAGAGGACCCAACAT	PCR with ID401 verifies insertion in X-2-DW
ID903-X-3-up-out	TGACGAATCGTTAGGCACAG	PCR with ID339 verifies insertion in X-3-UP
ID904-X-3-down-out	CCGTGCAATACCAAAATCG	PCR with ID401 verifies insertion in X-3-DW
ID905-X-4-up-out	CTCACAAAGGGACGAATCCT	PCR with ID339 verifies insertion in X-4-UP
ID906-X-4-down-out	GACGGTACGTTGACCAGAG	PCR with ID401 verifies insertion in X-4-DW
ID339-TEF1_test_rv	GCTCATTAGAAAGAAAGCATAGC	Verification of insertion of constructs containing TEF1

Table 3.2. List of plasmids used in this study.

Name	Description	Reference
Integrative plasmid set with URA3 selection marker flanked with direct repeats		
pCfB126	pX-2-USER-URA3-DR	[18]
pCfB127	pX-3-USER-URA3-DR	[18]
pCfB128	pX-4-USER-URA3-DR	[18]
pCfB383	pXI-1-USER-URA3-DR	[18]
pCfB384	pXI-2-USER-URA3-DR	[18]
pCfB385	pXI-3-USER-URA3-DR	[18]
pCfB387	pXI-5-USER-URA3-DR	[18]
pCfB129	pXII-1-USER-URA3-DR	[18]
pCfB120	pXII-2-USER-URA3-DR	[18]
pCfB130	pXII-4-USER-URA3-DR	[18]
pCfB131	pXII-5-USER-URA3-DR	[18]
pCfB49	pXII-1-ccdB-USER-URA3-DR	[18]
Plasmids that contain LoxP-flanked selection marker cassettes		
pUG6	LoxP-KanMX	[19]
pUG27	LoxP-SpHIS5	[19]
pUG72	LoxP-KlURA3	[19]
pUG73	LoxP-KlLEU2	[19]
pSA40	LoxP-CaLYS5	[23]
Episomal replication vectors with USER cassette (derived from pESC, Agilent)		
pCfB54	pESC-URA-ccdB-USER	This study
pCfB55	pESC-HIS-ccdB-USER	This study
pCfB132	pESC-URA-USER	This study
pCfB291	pESC-HIS-USER	This study
pCfB220	pESC-LEU-USER	This study
EasyClone integrative vector set with loxP-flanked selection markers		
pCfB255	pX-2-LoxP-KlURA3	This study
pCfB353	pX-2-LoxP-KanMX	This study
pCfB257	pX-3-LoxP-KlLEU2	This study
pCfB258	pX-4-LoxP-SpHiS5	This study
pCfB388	pXI-1-LoxP-KlLEU2	This study
pCfB389	pXI-2-LoxP-KlURA3	This study
pCfB390	pXI-3-LoxP-KlURA3	This study
pCfB391	pXI-5-LoxP-SpHIS5	This study
pCfB259	pXII-1-LoxP-KlLEU2	This study
pCfB260	pXII-2-LoxP-CaLYS5	This study
pCfB262	pXII-4-LoxP-SpHIS5	This study
pCfB261	pXII-5-LoxP-SpHIS5	This study
Plasmids containing genes encoding fluorescent proteins		
pWJ1163	CFP	[24]
pWJ1165	YFP	[24]
pWJ1350	RFP	[25]
EasyClone vectors for expression of genes for fluorescent proteins in S. cerevisiae		
pCfB393	pX-2-LoxP-KlURA3-TEF1::CFP	This study
pCfB394	pX-3-LoxP-KlLEU2-TEF1::RFP	This study
pCfB395	pX-4- LoxP-SpHiS5-TEF1::YFP	This study
Episomal vectors for expression of genes for fluorescent proteins in S. cerevisiae		
pCfB396	pESC-URA-USER-TEF1::CFP	This study
pCfB397	pESC-LEU-USER-TEF1::RFP	This study
pCfB398	pESC-HIS-USER-TEF1::YFP	This study
Plasmid contains a cassette for expression of GFP in E. coli (for positive control in USER cloning)		
pCfB774	pmExpCtrl	Dr. Hao Lao, DTU

Flow cytometry analysis

Transformants were grown O/N in SC-His-Leu-Ura and 50 µL were used to inoculate 3 mL Delft medium in 24 deep well plates, where the cells were grown at 30°C with 300 rpm agitation. When the cultures had reached mid exponential phase they were harvested and fixed with paraformaldehyde according to the following protocol. Samples (1.5 mL) were taken and immediately cooled in ice-water bath and subsequently centrifuged at 4°C, 2000 x g for 2 min. Supernatant was removed and pellet was resuspended in 200 µL of 2 % paraformaldehyde. The mix was incubated on ice for 1 hour and subsequently centrifuged at 4°C, 2000 x g for 2 min. Finally, the paraformaldehyde was removed and pellet was resuspended in 200 µl PBS buffer. The fixed cells were stored at 4°C until FACS analysis (max 1-2 days).

Cells were analyzed on a BD FACSAria equipped with three solid state diode lasers: air-cooled Coherent™ Sapphire™ solid-state diode laser (488 nm, 100 mW), air-cooled Coherent™ Yellow Green laser (561 nm, 100 mW), and an air-cooled Coherent™ Deep Blue laser (445 nm, 50 mW). The following filters were used: FITC-A, PE-Cy5-A and mCFP-A for analysis of emission from YFP, RFP and CFP, respectively. Compensation was performed according to manufacturer's protocol (BD FACSAria II User's Guide).

Flow cytometry data sets were analyzed and interpreted by software packages derived from the open source platform of Bioconductor [26].Outliers were removed by pre-gating on FSC and SSC data sets with the rule for outliers set at 90% quantile region. Cells were analyzed for their mean values, extracted as vectors and plotted by the *scatterplot3d* function (Ligges & Maechler).

Cre-Lox mediated selection marker loop out

Strains were transformed with pSH65 (EUROSCARF) harboring the *cre* gene under control of the *GAL10* promoter, and transformants were selected on YPD containing 10 µg mL^{-1} phleomycin (InvivoGen). Single colonies were picked and grown in YPD for 4-6 hours, harvested by centrifugation and resuspended in YPGal, where they were subsequently grown for another 12-16 hours. Dilutions of the culture were then plated on YPD plates and the emerging colonies were replica-plated on YPD, SC-Ura, SC-His and SC-Leu to verify that all three markers had been looped out. Strains showing successful triple selection marker loop-out were analyzed by flow cytometry as described above using SC-complete media as growth media.

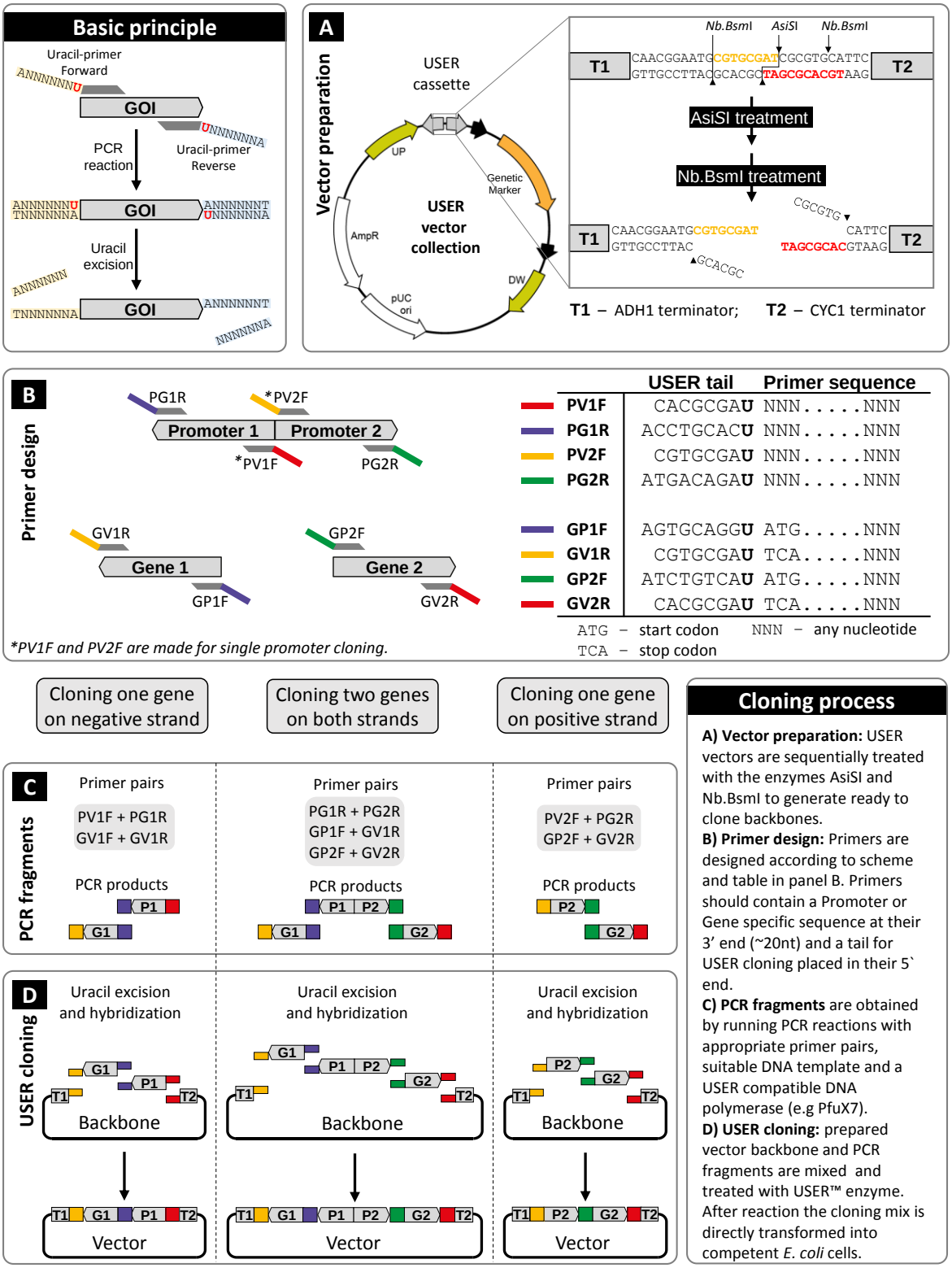


Figure 3.1. Overview of the procedure for cloning genes into EasyClone vectors. Detailed protocol can be found as supplementary material (**Figure S 3.1**).

Results and discussion

In order to decrease the turnaround time in the metabolic engineering cycle, two sets of plasmids, one episomal and one integrative were created (**Figure 3.2**). The episomal set was derived from a subset of the pESC plasmid series (Agilent, USA). Specifically, the multiple cloning sites and *GAL1/GLA10* promoters were were replaced by uracil excision based cloning cassette, AsiSI/Nb.BsmI [27], hence, making it USER cloning and USER fusion compatible, see **Figure 3.1** [17], [21]. The integrative vector set, which we named EasyClone, is based on the integrative plasmids from [18]. Specifically, we chose the vectors in the set, where matching integration sites were shown to accept foreign DNA without affecting fitness of the strain and where gene expression was high [18].For these vectors For these vectors the *K. lactis URA3* selection cassette was substituted for one of five different selective markers (see **Figure 3.2**). In order to be able to reuse the introduced selection markers the different markers are all flanked by *LoxP* sites, whereby the selection marker can be looped out by Cre recombinase mediated recombination [19].

Both episomal and integrative plasmids contain two terminator sequences in opposite directions flanking the USER cloning cassette. This facilitates incorporation of two genes and a bi-directional promoter, while the option of incorporating only one single gene with one-directional promoter remains (**Figure 3.1**). The design of the cloning cassette ensures directional cloning. It also provides flexibility for the combination of different genes with different promoters using the same gene PCR fragment for any combination as long as the position of the gene is maintained, i.e. Gene1 or Gene2. The different promoter fragments can be combined with any genes having the specified eight-nucleotide overhang, which allows for high-throughput cloning in a combinatorial setup.

In order to create a proof-of-concept for the plasmid set, an experiment was set up where the expression of three genes encoding three different fluorescent proteins from either episomal plasmids or from three integration sites in the genome was tested and compared (**Figure 3.3**). *CFP*, *YFP* and *RFP* were cloned into pESC-USER and three integration- plasmids, whereby six plasmids were constructed: pESC-CFP-URA, pESC-RFP-LEU, pESC-YFP-HIS, pX-2-CFP-LoxP-URA, pX-3-RFP-LoxP-LEU, and pX-4-YFP-LoxP-HIS. Strains were constructed harbouring either the three pESC-xFP plasmids or the three integration xFP expression fragments.

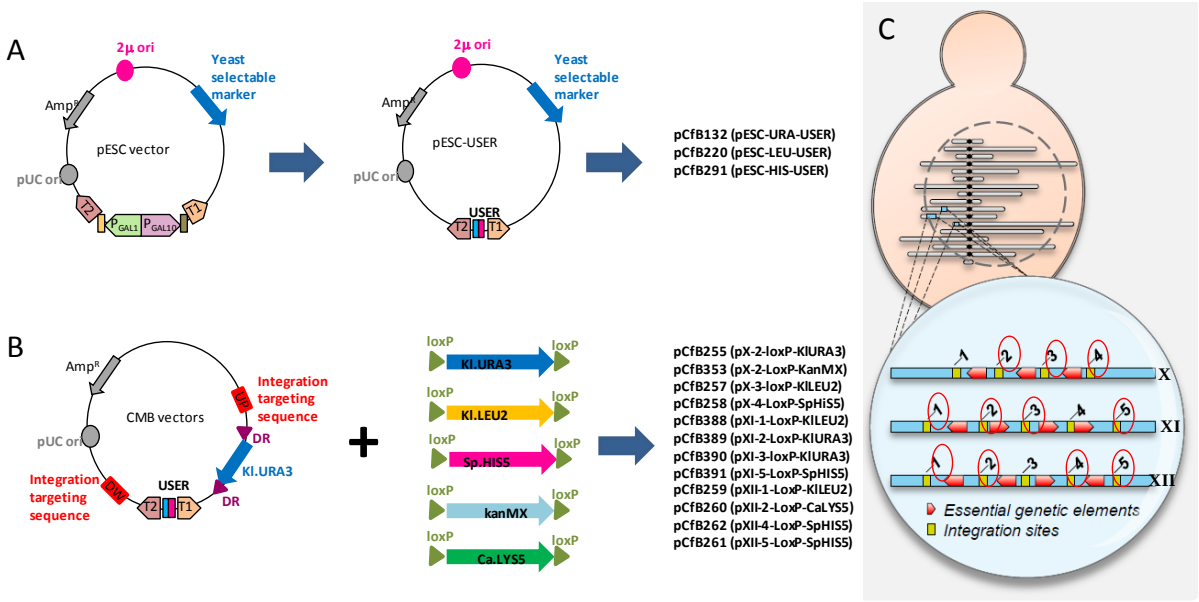


Figure 3.2. Plasmid construction. A, Episomal vectors are based on the pESC vector (Agilent, USA), where the multiple cloning sites and galactose-induced promoters were replaced by uracil excision based cloning cassette. B, Integrative vectors are based on the vectors described in Mikkelsen et al (2012). The *URA3* selection cassette flanked by direct repeats was exchanged with one of five different selective markers as indicated. The five markers are all flanked by *LoxP* sites allowing Cre mediated marker loop out. C, integration sites are organized in clusters on chromosomes X, XI and XII as presented by Mikkelsen at al. All integration sites (yellow boxes) are separated by either genetic elements which are essential for growth or by genes essential for maintaining wild type growth rates (red boxes). Numbers encircled in red represent Integration sites that are covered by the vector set presented here.

The three integration fragments carrying the three fluorescent protein encoding genes were transformed into yeast in one single transformation event. From this triple transformation 16 clones were tested for correct insertion by PCR. For seven clones all the expected bands were seen on DNA electrophoresis and all of these exhibited triple fluorescence from CFP, RFP and YFP (results not shown). This showed that it is indeed possible to do triple integration in a targeted fashion with a relatively high success rate (44 %).

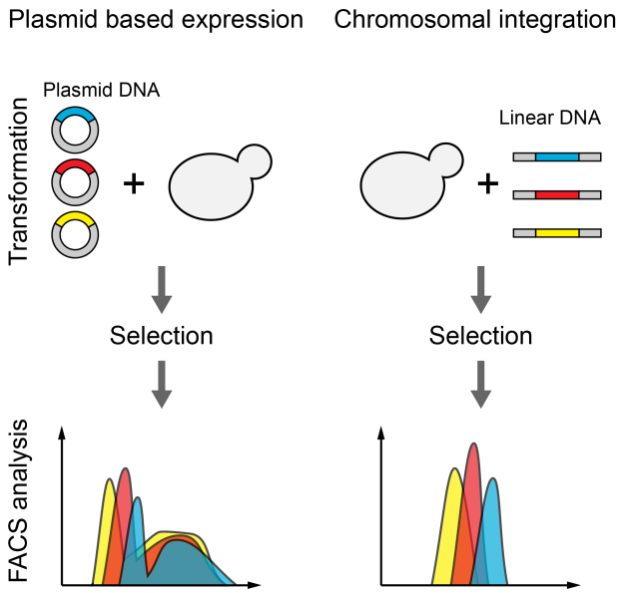


Figure 3.3. Experimental setup. *CFP*, *YFP*, and *RFP* were cloned into either episomal or integration vectors under control of the strong *TEF1* promoter. *S. cerevisiae* was transformed with either three episomal or three integration vectors followed by flow cytometric analysis for presence of the three fluorescent proteins.

In order to test the individual production of the three fluorescent proteins in the two different strains containing the genes either on episomal 2µ plasmids or as a triple genomic integration, the fluorescence levels of single cells were analyzed by flow cytometry (**Figure 3.4**). Triple fluorescent protein production in strains containing the relevant genes as genomic integrations was much more uniform as compared to strains were they were harbored on episomal plasmids. The mean levels of fluorescence were in the same range for the two expression systems, whereas the standard deviations for cells expressing the three fluorescent proteins from episomal plasmids were 4-5 times larger than for cells expressing from triple integrations (**Table 3.3**).

Table 3.3. Log_{10} mean values with standard deviations for each fluorescence signal for cells expressing CFP, RFP, and YFP.

	Integration	**Episomal plasmids**
CFP	3. 11 ± 0.21	3.4 ± 0.83
RFP	3.1 ± 0.2	3.3 ± 0.98
YFP	3.38 ± 0.22	3.41 ± 0.94

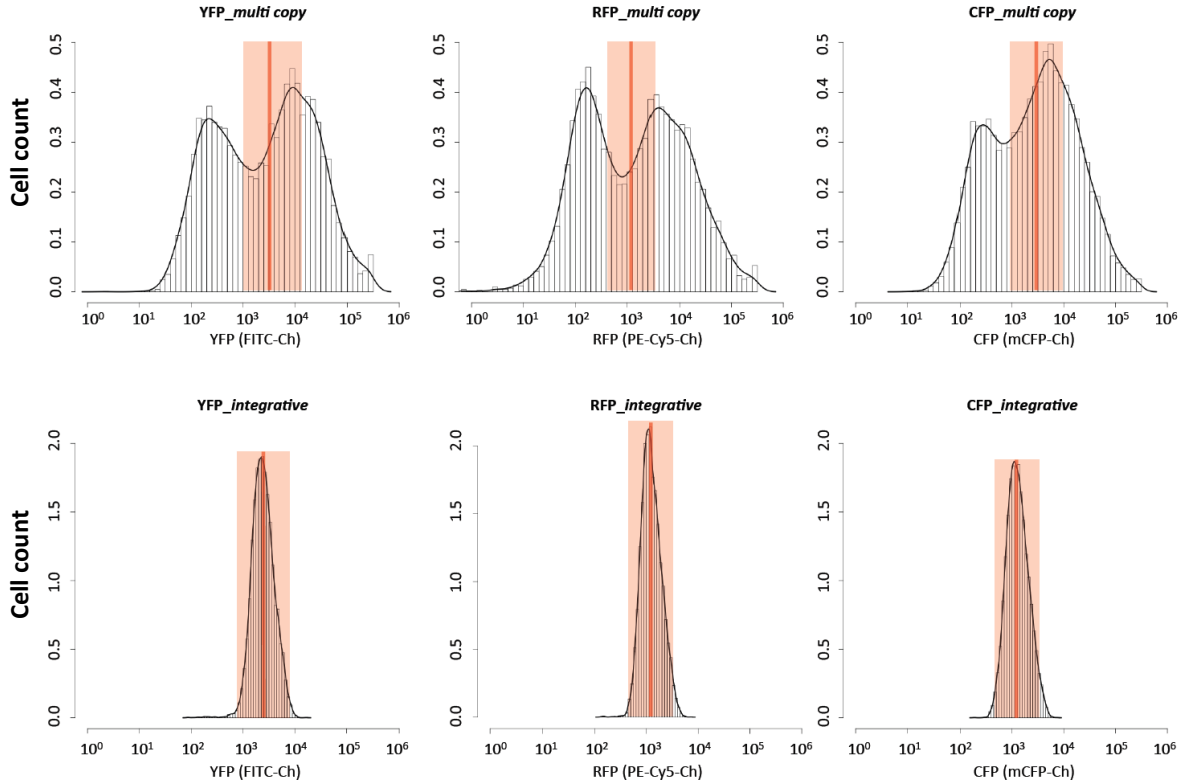


Figure 3.4. Flow cytometry on *S. cerevisiae* strains co-expressing *YFP, RFP* and *CFP* from either three episomal plasmids (top panel) or from triple integrations on the genome (bottom panel). The Log_{10} mean value ± 15% for each color is indicated with a deep red vertical line and a light red shading, respectively.

Next, we determined the levels of simultaneous production of the three fluorescent proteins in individual cells and plotted the data into three dimensional plot (representing levels of YFP, CFP, and RFP), see **Figure 3.5**. This analysis convincingly demonstrated that cells expressing the three genes from episomal plasmids are much more scattered throughout the whole three-dimensional space, whereas the cells with genomic integrations are in a much more defined space. As a measure for uniformity of protein production in the two systems we defined that cells containing a fluorescent signal deviating from Log_{10} mean ± 15% for each color are identical for all three colors (highlighted in red on **Figure 3.5**). Based on this definition, only 6% of the cells harboring the episomal expression system contained identical levels of fluorescent proteins. In contrast, more than 95 % of the cells were identical when the genes were integrated into the genome. This clearly demonstrates the advantage of the EasyClone plasmid set for the construction of complex pathways in yeast, as it is important to have stable and concomitant expression of all genes introduced in each cell in order to draw sensible conclusions.

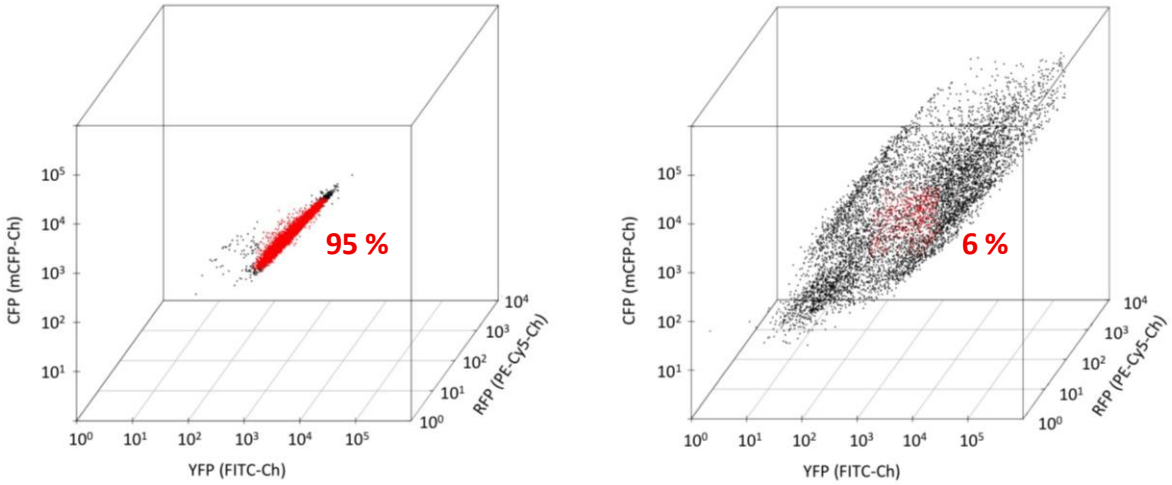


Figure 3.5. 3D plot of the fluorescence levels of cells expressing CFP, RFP and YFP from either triple genomic integrations (left box) or from episomal plasmids (right box). Red dots represent cells having fluorescence intensities for all three fluorophores being within Log_{10} mean ± 15% for each color and black dots represent cells with one or several fluorescence levels being outside of mentioned interval.

For construction of large pathways or for repeated cycles of metabolic engineering it is important that all markers used in a multiple integration experiment can be recycled. We therefore tested whether it would be possible to eliminate all three selection markers used for the integration of the genes encoding CFP, RFP, and YFP simultaneously. A strain expressing all three fluorescent proteins was transformed with a *cre*-containing plasmid and *cre* was subsequently induced by growing the transformants on galactose to allow for production of Cre recombinase. 96 clones generated in this manner were tested for successful selection marker loop out and eight of these showed a histidine, leucine and uracil auxotrophy. All eight strains were re-tested for fluorescence and all showed fluorescence patterns, which were indistinguishable from the pattern produced by the parent strain (**Figure S 3.2**). The low level of ura his leu clones was most likely due to the proximity of the integration sites. The three integration sites were all on the same chromosome, which meant that there were 6 *LoxP* sites introduced within a fairly small genomic region of 42 kb. Hence, there was a risk of recombination between LoxP sites in two different integration sites with a lethal loss of an essential gene element to follow. Indeed we obtained efficiencies above 90% for removal of selection markers integrated on different chromosomes (unpublished results).

Conclusions

In conclusion, we have shown that using our vector set it is possible to introduce up to three integration cassettes in *S. cerevisiae* genome simultaneously. Each integration cassette can be constructed to carry 1-2 genes. The selection markers used for the integration can be looped-out simultaneously without the loss of the integrated genes. We also show that expression of multiple genes from integrative cassettes leads to more homogeneous expression within the yeast population than expression from multiple episomal vectors. Combined with the fact that vector construction is based on highly efficient USER cloning, our system is well-suited for the construction of cell factories containing multiple genetic modifications, even in a high throughput manner, see **Figure 3.6**. The plasmids are freely available on request.

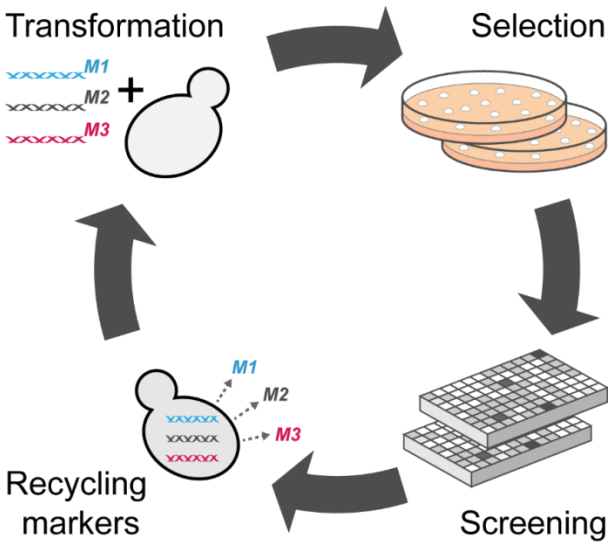


Figure 3.6. Illustration of how our method can be integrated into efficient strain construction in metabolic engineering. M1, M2 and M3 represent selective markers associated with genetic elements used for yeast transformation.

References

[1] I.-K. Kim, A. Roldão, V. Siewers, and J. Nielsen, “A systems-level approach for metabolic engineering of yeast cell factories.,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 228–48, Mar. 2012.

[2] N. a Da Silva and S. Srikrishnan, “Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae.,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 197–214, Mar. 2012.

[3] A. Krivoruchko, V. Siewers, and J. Nielsen, “Opportunities for yeast metabolic engineering: Lessons from synthetic biology.,” *Biotechnol. J.*, vol. 6, no. 3, pp. 262–76, Mar. 2011.

[4] M. S. Siddiqui, K. Thodey, I. Trenchard, and C. D. Smolke, “Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools.,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 144–70, Mar. 2012.

[5] K.-K. Hong and J. Nielsen, “Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries.,” *Cell. Mol. Life Sci.*, vol. 69, no. 16, pp. 2671–90, Aug. 2012.

[6] a B. Futcher, “Copy number amplification of the 2 micron circle plasmid of Saccharomyces cerevisiae.,” *J. Theor. Biol.*, vol. 119, no. 2, pp. 197–204, Mar. 1986.

[7] D. J. Mead, D. C. J. Gardner, and S. G. Oliver, “The yeast 2 μ plasmid: strategies for the survival of a selfish DNA,” *Mol. Gen. Genet. MGG*, vol. 205, no. 3, pp. 417–421, 1986.

[8] S. Alberti, A. A. D. Gitler, and S. Lindquist, “A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae,” *Yeast*, no. June, pp. 913–919, 2007.

[9] R. D. Gietz and S. Akio, “New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites,” *Gene*, vol. 74, no. 2, pp. 527–534, Dec. 1988.

[10] I. Sadowski, T.-C. Su, and J. Parent, “Disintegrator vectors for single-copy yeast chromosomal integration,” *Yeast*, vol. 24, no. 5, pp. 447–455, 2007.

[11] R. S. Sikorski and P. Hieter, “A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in,” *Genetics*, no. 122, pp. 19–27, 1989.

[12] F. W. Lee and N. a Da Silva, “Improved efficiency and stability of multiple cloned gene insertions at the delta sequences of Saccharomyces cerevisiae.,” *Appl. Microbiol. Biotechnol.*, vol. 48, no. 3, pp. 339–45, Sep. 1997.

[13] X. Wang, Z. Wang, and N. a Da Silva, “G418 Selection and stability of cloned genes integrated at chromosomal delta sequences of Saccharomyces cerevisiae.,” *Biotechnol. Bioeng.*, vol. 49, no. 1, pp. 45–51, Jan. 1996.

[14] J. L. Hartley, G. F. Temple, and M. A. Brasch, “DNA Cloning Using In Vitro Site-Specific Recombination,” *Genome Res.*, vol. 10, no. 11, pp. 1788–1795, Nov. 2000.

[15] D. D. G. Gibson, L. Young, R. R.-Y. Chuang, J. C. Venter, C. A. 3rd Hutchison, and H. O. Smith, “Enzymatic assembly of DNA molecules up to several hundred kilobases,” *Nat. Methods*, vol. 6, no. 5, pp. 343–345, 2009.

[16] C. Engler, R. Kandzia, and S. Marillonnet, “A one pot, one step, precision cloning method with high throughput capability.,” *PLoS One*, vol. 3, no. 11, p. e3647, Jan. 2008.

[17] H. H. Nour-Eldin, B. G. Hansen, M. H. H. Nørholm, J. K. Jensen, and B. a Halkier, “Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments.,” *Nucleic Acids Res.*, vol. 34, no. 18, p. e122, Jan. 2006.

[18] M. D. Mikkelsen, L. D. Buron, B. Salomonsen, C. E. Olsen, B. G. Hansen, U. H. Mortensen, and B. A. Halkier, “Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform.,” *Metab. Eng.*, vol. 14, no. 2, pp. 104–11, Mar. 2012.

[19] U. Gueldener, J. Heinisch, G. J. Koehler, D. Voss, and J. H. Hegemann, “A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.,” *Nucleic Acids Res.*, vol. 30, no. 6, p. e23, Mar. 2002.

[20] M. H. H. Nørholm, “A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering,” *BMC Biotechnol.*, vol. 10, p. 21, 2010.

[21] F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, and B. a Halkier, “USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.,” *Nucleic Acids Res.*, vol. 35, no. 7, p. e55, Jan. 2007.

[22] R. D. Gietz and R. H. Schiestl, “Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method,” *Nat. Protoc.*, vol. 2, no. 1, pp. 35–37, 2007.

[23] S. Ito-Harashima and J. H. McCusker, “Positive and negative selection LYS5MX gene replacement cassettes for use in Saccharomyces cerevisiae.,” *Yeast*, vol. 21, no. 1, pp. 53–61, Jan. 2004.

[24] R. J. D. R. Reid, M. Lisby, and R. Rothstein, “Cloning-free genome alterations in saccharomyces cereuisiae using adaptamer-mediated PCR,” *Methods Enzymol.*, vol. 350, no. 1983, pp. 258–277, 2002.

[25] M. Lisby, U. H. Mortensen, and R. Rothstein, “Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre.,” *Nat. Cell Biol.*, vol. 5, no. 6, pp. 572–7, Jun. 2003.

[26] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang, “Bioconductor: open software development for computational biology and bioinformatics.,” *Genome Biol.*, vol. 5, no. 10, p. R80, Jan. 2004.

[27] B. G. Hansen, B. Salomonsen, M. T. Nielsen, J. B. Nielsen, N. B. Hansen, K. F. Nielsen, T. B. Regueira, J. Nielsen, K. R. Patil, and U. H. Mortensen, “Versatile enzyme expression and characterization system for Aspergillus nidulans, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case.,” *Appl. Environ. Microbiol.*, vol. 77, no. 9, pp. 3044–51, May 2011.

Supplementary Information

Contents

List of Supplementary Figures:

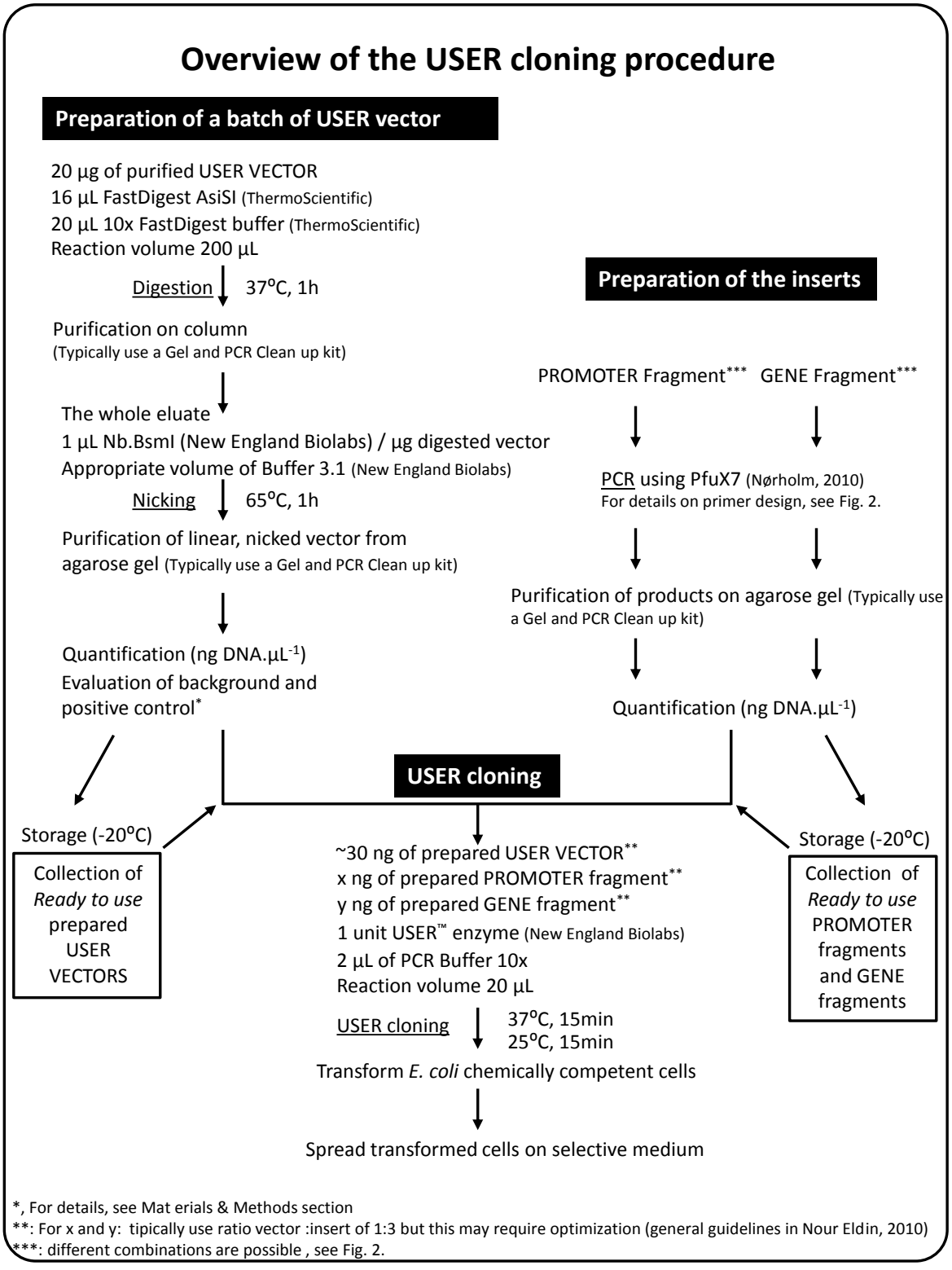


Figure S 3.1. USER cloning protocol

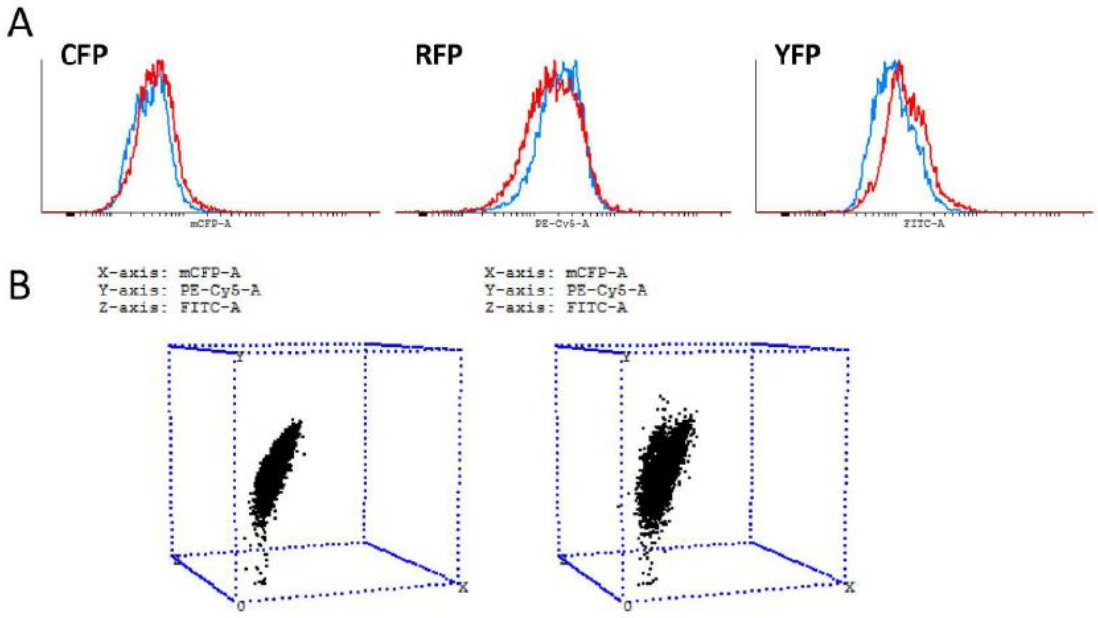


Figure S 3.2. FACS analysis of the cell before and after the triple selection marker loop-out. A, Single color overlay histogram, for cyan (CFP), red (RFP) and yellow (YFP) fluorescent proteins signals before (blue) and after (red) marker loop-out. B, 3D plots of CFP, RFP and YFP before (left) and after (right) marker loop-out.

Chapter 4

Gene amplification in *Saccharomyces cerevisiae*

Title[3]

Controlled gene amplification for high, tunable, selection-free gene expression in the yeast *Saccharomyces cerevisiae*

Abstract

Introduction of multi enzyme pathways in to yeast requires precise control over the heterologous gene expression levels. Current methods (i.e., episomal multicopy plasmids also known as 2µ plasmids) allow high gene expression levels to be reached, but do not ensure genetic stability if a proper selection pressure is not applied. The following chapter describes the design and validation of a gene amplification system that enables stable and selection-free gene expression in the yeast *Saccharomyces cerevisiae*. Here, the collection of special yeast strains have been developed with the capability of autonomous gene amplification up to ten copies. As a proof of concept two genes encoding red and cyan fluorescent proteins were amplified using the developed method and the fluorescence intensities were compared to the ones obtained by a plasmid based system.

Most notably, the gene amplification method was successfully applied in the subsequent study described in the **Chapter 5**, where it was used for alleviation of metabolic bottlenecks in the heterologous pathway for vanillin-glucoside production in *S. cerevisiae*.

[3] Manuscript in preparation

Introduction

A major challenge when engineering cell factories for industrial production is to achieve high-gene expression levels without compromising genetic stability. In the popular expression hosts *Saccharomyces cerevisiae* and *Escherichia coli*, high copy numbers are readily achieved by expressing genes from multi copy plasmids. But plasmids are gradually lost when the cell divides because plasmids are not as stably inherited as chromosomes [2]. The instability associated with plasmid based expression is a major cause of productivity loss in long-term cultivations [3]–[6]. On the other hand genomic integration provides ordered genetic inheritance and thereby increases the genetic stability of heterologous genes. However, to meet plasmid based expression levels, it is necessary to integrate multiple copies of each gene. This can be achieved by the time consuming process of repetitive rounds of genetic transformation. To solve the problems of combining high copy numbers with fast strain construction, different methods that allow multi-copy strains to be constructed in a few steps have been developed. Generally such methods involve either targeting of natural repeats whereby multiple gene copies can be integrated at once or it involves integration of a single gene copy followed by gene amplification. By targeting natural repeats such as ribosomal DNA [7] or δ-sequences [8] in *S. cerevisiae*, strains estimated to contain up to 300-400 gene copies were obtained and expression levels equaled those that can be achieved by multi copy plasmids at least in some cases [7]. But strains were generally unstable, because the method results in tandem arrangement of the gene copies even when the targeted repeats are dispersed throughout the genome [9]–[11]. In general, the copy number of DNA repeats positioned in close proximity is unstable, because repeated elements are easily lost through direct repeat recombination.

Alternatively, high copy strains can be achieved through gene amplification. In *E. coli,* strains that yielded up to 40 integrated gene copies were achieved by coupling gene amplification to antibiotic resistance [12]. This was achieved by placing the genes to be amplified and an antibiotic resistance gene in between repetitive sequences. When the cells divide, recombination errors will in some cases result in a duplication of the repeat flanked region. As duplication always results in an extra copy of the gene(s) of interest as well as an extra copy of the resistance gene, duplications can be selected for by growing strains in media with increasing concentrations of the relevant antibiotic. When the operons responsible for lycopene and poly-hydroxybutyrate production were amplified using this method, the yield exceeded that of strains which expressed the same operons from multi copy plasmids by 1.6-4 fold [12]. But similarly to the methods that target repeats, the method results in strains where the amplified genes are arranged in tandem and according to the authors it was necessary to knock out the recombination protein *recA* to stabilize strains after amplification. If the strain is to be used as a cell factory, the knock-out of enzymes essential for recombination is

disadvantageous because such deletions generally results in strains that are less robust and grow slower [12]. Interestingly, a gene amplification method where genes were not arranged in tandem has also been developed. By exploiting the natural mechanism of transposon migration Boeke and co-workers successfully amplified three genes and obtained strains with up to ten integrated gene copies [13]. Strains generated by transposon mediated gene amplification are presumably stable, as gene copies are dispersed throughout the genome, but the amplification process occurs through the error prone reverse transcriptase which means that the amplified gene copies are likely to contain mutations. Moreover there is limitation associated with the method, as it was observed that the amplification rate and copy number decreased substantially when the size of the DNA to be amplified was increased from 1kb to 4.6kb [14].

Here we describe an inducible, plasmid free gene expression system that enables fast construction of stable strains with high copy numbers in the yeast *S. cerevisiae*. Using our gene amplification system, genes are copied and integrated at well-defined positions in the genome. More specifically, genes are inserted in non-coding intergenic regions where it has previously been shown that integration does not influence fitness of the strain by adversely affecting neighboring genes [15]. Moreover, these integrations sites are separated by essential genes which prevent the loss of amplified genes through direct repeat recombination. This should ensure long-term strain stability and provide selection free gene expression. Gene amplification is achieved through double strand break stimulated gene conversion. Copy numbers can be precisely controlled and are determined by the number of gene amplification cassettes present in the pre-engineered strain. We demonstrate that strains with up to ten integrated copies can be constructed and that strains generated using our system remain stable in long-term cultivations.

Results

Design and construction of the gene amplification system

To develop a plasmid free gene expression system that enables construction of stable strains with high copy numbers, we set out to develop a gene amplification system where gene copies are integrated at well-defined positions in the genome. The system utilizes an endonuclease to stimulate homologous recombination at specific loci. Gene amplification is carried out in a strain that has been pre-engineered to contain a number of gene amplification cassettes. Each gene amplification cassette contains an I-SceI cleavage site [16] where specific double strand breaks can be made. Moreover, each cassette contains a counter-selectable marker, which in our system was the *URA3* gene of *Kluyveromyces lactis* flanked by two recombination sequences denoted A and B. The recombination sequences A and B enable gene amplification by homologous recombination upon I-SceI expression (see **Figure 4.1 A** for a schematic representation of the process). In this case the recombination sequences A and B where approximately 500 and 1000 base pairs in length, respectively, and were derived from genomic DNA of *Neurospora crassa* to avoid any similarity to the genomic DNA of *S.cerevisiae*.

To set up the gene amplification system in *S. cerevisiae* we constructed strains with one to ten amplification cassettes integrated on chromosome X, XI and XII at specific sites that were previously characterized by Mikkelsen et al [15] (**Figure 4.1**B). Integrations were made sequentially in three strains in parallel. This resulted in three types of GA strains; two with three amplification cassettes integrated on either chromosome X or XI and one with four amplification cassettes integrated on chromosome XII. Subsequently, strains with more than four amplification cassettes were generated by crossing the strains containing three or four gene GA cassettes with each other (e.g. a strain with seven amplification cassettes was constructed by crossing the strain with three cassettes on chromosome X with a strain with four cassettes on chromosome XII). For the general notion gene amplification strains were named GAx followed by number, e.g. GAx10 is the strain with ten amplification cassettes. All strains were made in minimum two variants of which one had mating type "**a**" and the other had mating "**α**". This allows construction of diploid strains with up to 20 gene copies by mating. In order to easily integrate and amplify any gene or pathway of interest in the GAX strains, a USER cloning vector (pCSN) was constructed. pCSN contains a uracil excision based cloning cassette AsiSI/Nb.BsmI cassette [17], flanked by two yeast terminators of *ADH1* and *TRP1*, a selective marker and the recombination sequences A and B targeting integration into an amplification cassette (**Figure 4.5 A**).

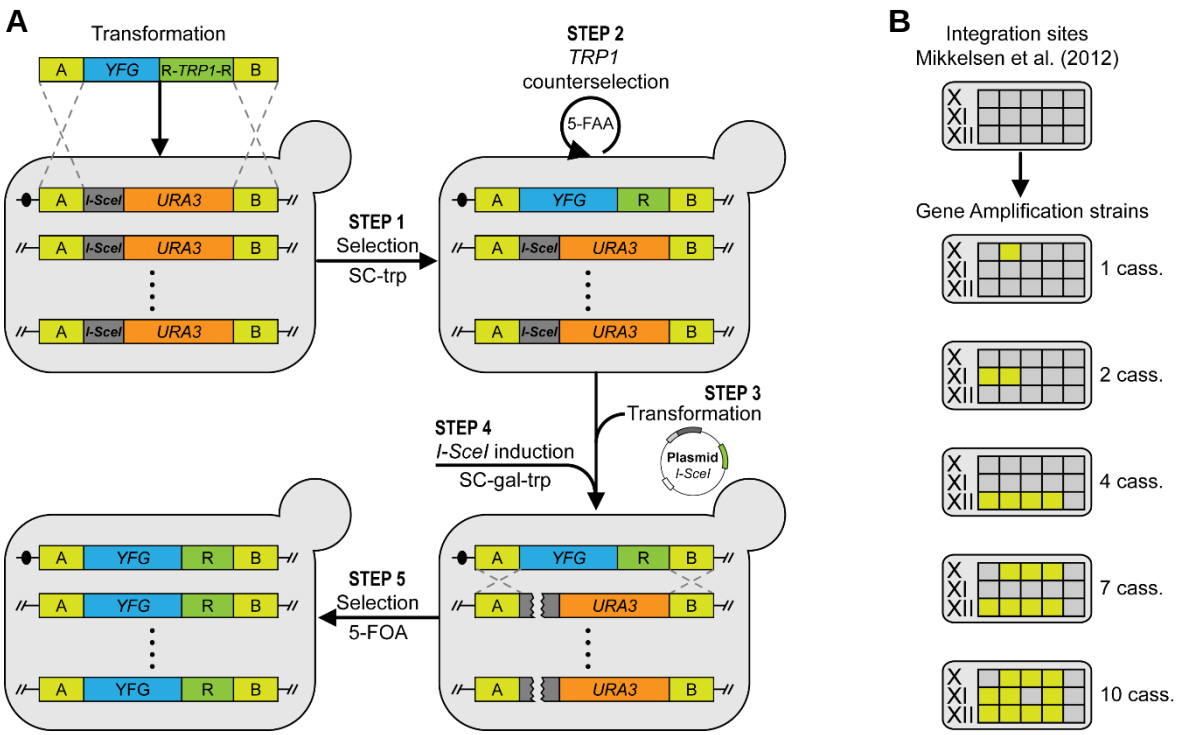


Figure 4.1. The gene amplification (GA) system. **A**) General scheme of the amplification process. **Step 1**: A strain which has been pre-engineered to contain a number of amplification cassettes is transformed with a substrate containing "your favorite gene/s" (*YFG*) to be amplified, a counter selectable marker flanked by direct repeats (*R-TRP1-R*) and the recombination sequences A and B. **Step 2**: The *TRP1* marker is recycled by plating cells on counter-selective media (SC-5-FAA). **Steps 3** and **4**: I-SceI is introduced to the cell on a plasmid and its expression is induced by transferring cells to induction media (SCgal-trp). Expression of I-SceI results in double strand breaks in all amplification cassettes except the one where *YFG* has integrated. After induction, the breaks are repaired by homologous recombination which results in the displacement of *URA3* with *YFG*. **Step 5**: Cells where all amplification cassettes have been successfully replaced by *YFG* are identified by plating cells on *URA3* counter selective media (SC-5-FOA). **B**) Simplified representation of integration platform used for construction of the GA strains. Three rows represent defined chromosomes (X, XI and XII) and five columns correspond to the integration site number (1 to 5) for each chromosome [15]. Yellow box depicts the location of the GA cassette. This panel represents only a fraction of all available strains developed in this study.

A Proof-of-concept

For a proof of concept the GA system was employed for controlled overexpression of the genes encoding the two model proteins – cyan fluorescence protein (CFP) and red fluorescence protein (RFP). These two proteins were chosen as model proteins because their fluorescence intensity can easily be quantified using flow cytometry. Gene targeting fragment for integrating *CFP* and *RFP* in the GA strains were prepared by digesting the pCSN-CFP-RFP plasmid with NotI endonuclease. To construct strains with different copy numbers of *CFP* and *RFP*, the gene targeting substrate was transformed into strains containing one, two, four, seven and ten cassettes. Subsequently, gene amplification was achieved by transformation with the I-sceI encoding plasmid, pWJ1320-TRP, followed by a series of replica plating and selection events as described in the **Materials and Methods** section. In order to construct strains with higher than ten gene copies (up to 20) we constructed diploid cells by mating haploids of opposite mating type containing different copies of amplified reporter protein genes (**SI Table S 4.4**).

To demonstrate an application of the developed gene amplification system and to compare this approach to existing plasmid based systems (particularly, multi-copy 2µ plasmids), it was chosen to amplify two fluorescent proteins simultaneously in order to mimic a two-step biosynthetic pathway. For a two-step pathway, the successful production of the end-product requires the presence of both enzymes. From the results in Figure 4.2 it is evident that the GA system developed in this work is more advantageous compared to the plasmid based system in terms of simultaneous co-expression of two genes. This was true for all GA strains with different copy numbers of *CFP* and *RFP* genes. In all cases more than 90% of GA cells were fluorescing in both red and cyan colors. When these two proteins were expressed from two plasmids only 54% of the cells were fluorescing in both colors (**Table 4.1**). Moreover, fluorescence intensities of the strains expressing reporter genes from 2µ plasmids were scattered compared to the ones of the GA strains. This observation indicated that concentrations of CFP and RFP in the individual cells of 2µ strain varied significantly, i.e. a cell displaying high signal for RFP could have low or no signal for CFP and vice versa. To visualize the efficiency of co-production of two reporter proteins the RFP/CFP ratios have been plotted for each cell (see Figure 4.2C).

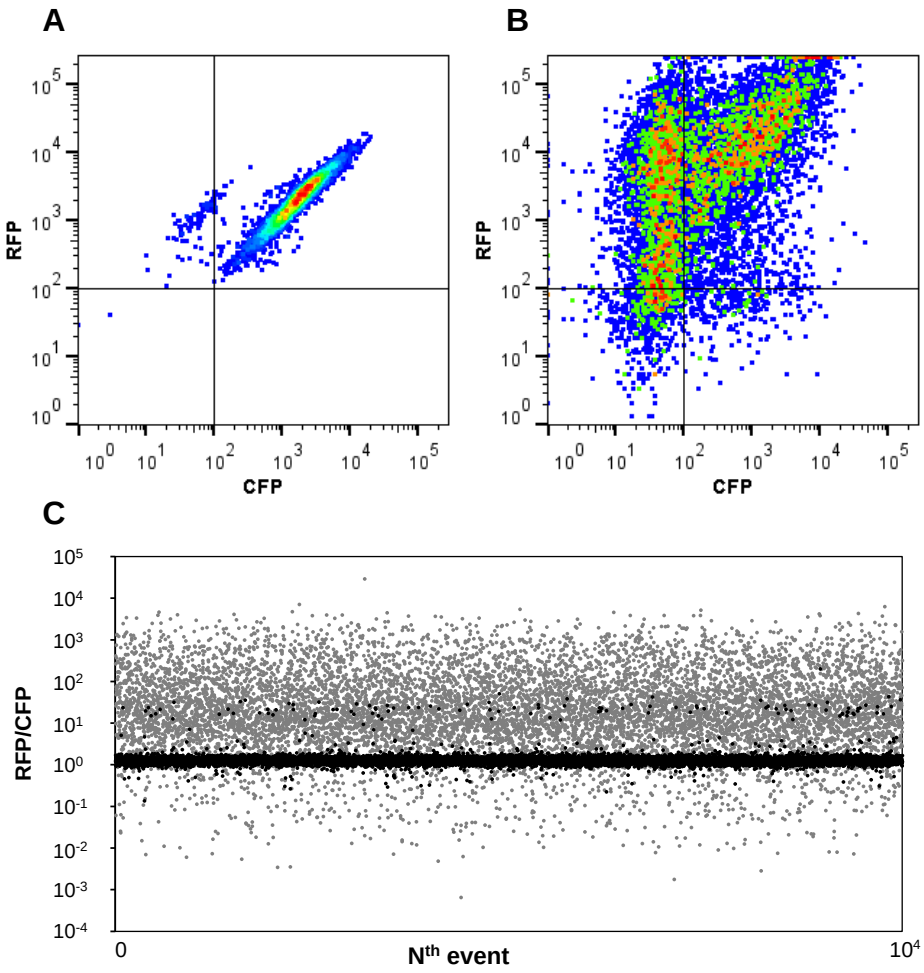


Figure 4.2. Scatter plots of CFP and RFP fluorescence levels observed in A) GAx10 based strains and B) in strains harboring two 2µ plasmids. C) The RFP/CFP fluorescence ratio in each individual cell represented for all analyzed strain population; (•) GAx10 and (•) 2µ plasmid strains. In all graphs CFP and RFP are plotted in Log10 scale.

Table 4.1. Average percentage of the cell population expressing reporter proteins. (S) – Cultivations were done in selective medium and (NS) – non selective medium.

Copies	% of all population			
	None	**CFP**	**RFP**	**Both**
0	93.6	1.7	3.5	1.2
1	17.8	10.9	9.4	61.9
2	3.0	4.1	2.9	90.0
4	0.2	0.4	0.7	98.7
7	0.1	0.1	2.2	97.6
10	0.1	0.0	0.7	99.2
2µ (S)	7.4	3.2	35.4	54.0
2µ (NS)	61.0	2.9	29.1	7.0

Next, it was investigated how the gene copy number reflects on the levels of produced fluorescent proteins and compared to the levels where the genes were expressed from 2µ plasmids. For this reason, mean fluorescence intensities (MFI) for two different colors (cyan and red) were measured in the two settings (i.e., GA and plasmid based). This analysis convincingly demonstrated that the fluorescence intensity for both colors increased linearly with the increasing gene copy number in the GA strains (see **Figure 4.3**). Not surprisingly, the same trend was also observed in the GA diploid strains. In addition, CFP production from the seven-copy GA strains exceeded the plasmid based expression system judging on MFI values. That was not true for RFP, where fluorescence levels were more than 60 times higher in the plasmid based strains compared to the highest levels obtained in the GA based diploid strain with 20 copies of *RFP* genes (**Figure 4.3**).

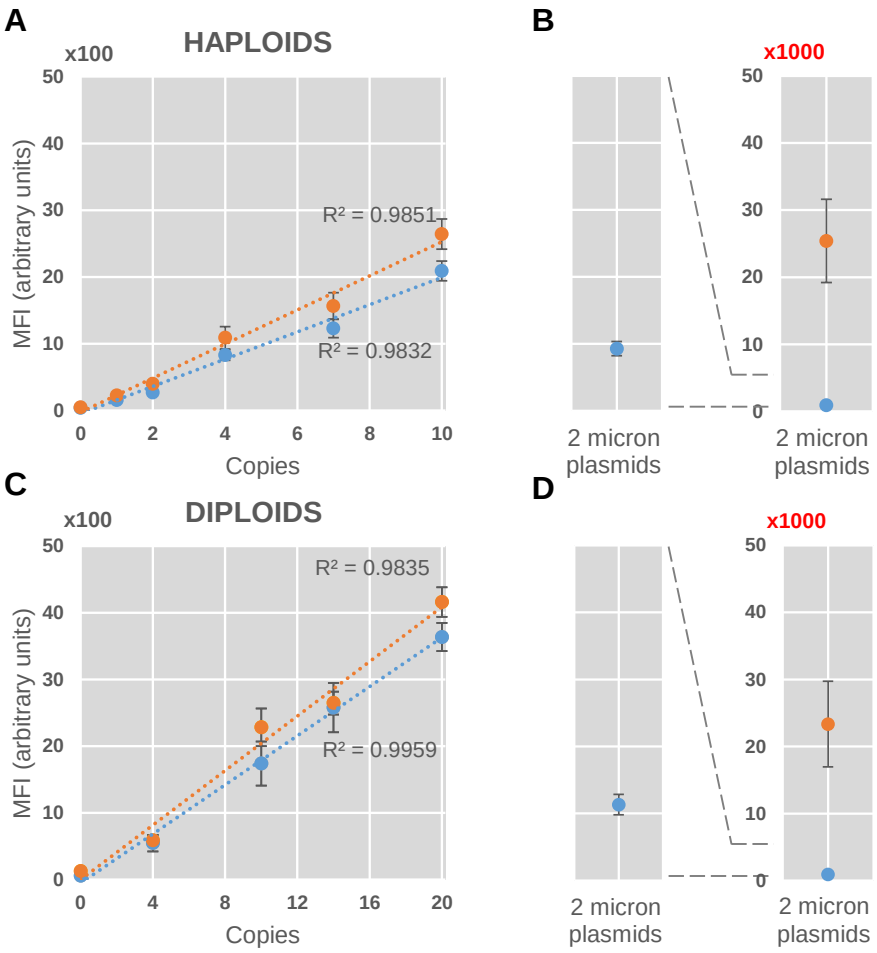


Figure 4.3. Mean fluorescence intensities (MFI) of *CFP* (•) and *RFP* (•) in GA and 2µ strains. MFIs observed in the haploid strains harboring up to ten copies of the integrated reporter genes (**A**) and in the strains where *CFP* and *RFP* were co-expressed on two independent 2µ plasmids (**B**). Correspondingly, MFIs in diploids strains with up to 20 copies (**C**) and plasmid based expression (**D**).

Genetic stability of the GA strains

One of the most important requirements of an industrial strain used as cell factories is the genetic stability in a long-term cultivations. To check this, the GAx10 and 2µ strains expressing *CFP* and *RFP* genes were cultivated for approx. 40 generations and samples for flow cytometry were taken at regular intervals. Two factors considering the stability of reporter genes were assessed; first, fluorescence levels of independent fluorophores throughout the cultivations and, second, the percentage of the cells that displayed fluorescence for both CFP and RFP. From the **Figure 4.4**A it can be seen that the GA strain with ten copies of *CFP* and *RFP* genes retained its initial fluorescence levels after 40 duplications, that was also true for the plasmid based strain. However, significant differences between the two strains were found when co-expression of the *CFP* and *RFP* genes was evaluated (see **Figure 4.4**B). Close to 100% of cells in all analyzed populations of GAx10 strain maintained fluorescence for both cyan and red colors throughout all experiment. Accordingly, only approx. 60 % of the cells expressing reporter genes from the 2µ based plasmids were found to produce both fluorescence proteins.

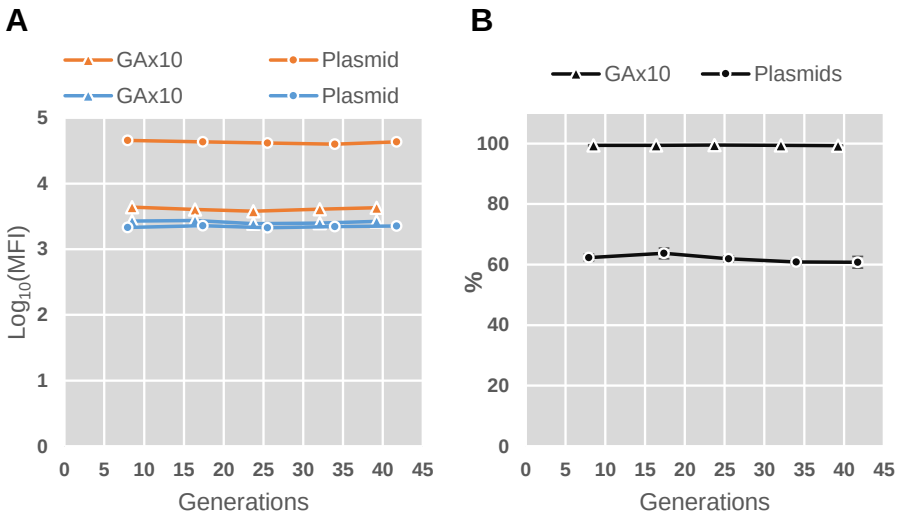


Figure 4.4. Genetic stability of the GA strains compared to the multi-copy plasmid based expression systems. **A**) Log_{10} MFI of the cells expressing *CFP* or *RFP* in two systems. **B**) Percentage of the cells that displayed fluorescence for both colors.

Discussion

In this work, a new method for gene amplification (GA) in yeast *S. cerevisiae* have been developed and validated. The GA system is conceptually different from the previously reported methods for construction of strains containing multi-copies of chromosomally integrated genes of interest [7], [8], [13]. Here, the amplification relies on a set of special made strains with a predefined **number** and **location** of GA cassettes located on three different chromosomes X, XI and XII. The increase in gene copy number occurs by the gene conversion mechanism [18] which is triggered by I-SceI induced DNA double strand breaks at the designated GA loci (see **Figure 4.1**).

Results of this study provide compelling evidence that the GA system enables amplification of DNA fragments of interest to a precise copy number (up to ten copies). This was confirmed when the system was used for overexpression of the two reporter proteins CFP and RFP. Mean fluorescence intensity increased linearly increased with the copy number of the *CFP* and *RFP* genes (**Figure 4.3**).To check if the higher gene copy number will result in adequate increase in protein production, diploid trains with various copy numbers (up to 20) of integrated *CFP* and *RFP* were constructed. Diploid data, however, must be treated with caution even though MFI values had increased with the similar trend as in haploids, the flow cytometry readings do not account for the increased mass and volume of a diploid cell. On the other hand, this indicated that it is possible to further improve protein production in haploids if GA strains with higher number of amplification cassettes were constructed.

One of the most important points of this work was the benchmarking of the GA based expression system to the 2μ plasmid based expression system. Several important criteria of these systems (based on the MFI of reporter proteins) were assessed; i.e., gene expression/protein production levels, production stability and homogeneity. These are key parameters of a successful industrially applicable cell factory. When MFI was compared for independent fluorophores, the levels of produced RFP in a strain harboring a 2μ based plasmid were orders of magnitude higher than in the GAx10 strain. In contrast, the CFP production of the GA strains was superior compared to plasmid based counterparts when the copy exceeded seven copies (**Figure 4.3**). The significantly different performance of the two plasmids was possibly caused by the uneven copy number of these plasmids. This can be explained by previously reported findings, where it was demonstrated that copy numbers of the 2μ plasmid can vary depending on the selection marker and the strength of the promoter used to express the gene of interest [19], [20]. Predominantly, beneficial properties of the GA system became evident when co-expression levels of the two reporter genes *CFP* and *RFP* in GA strains were compared to plasmid based counterparts (Figure 4.2 and **Table 4.1**). The ability to homogenously overexpress two genes in the GA strains is particularly advantageous in cases

where production of the compound of interest is depending on more than one enzyme. Lastly, the GA strains were demonstrated to be genetically stable in long-term cultivations for at least 40 generations (**Figure 4.4**). Strains with 10 copies of *CFP* and *RFP* retained initial fluorescence intensities as well as homogeneity after 40 generations in a synthetic complete medium without any selective pressure. The aptitude to maintain a stable number of genes is a result of an in-build "fail safe" system in GA strains which prevents the further cell propagation if any of amplified gene was lost through a direct repeat recombination [15]. Even though a selection pressure was constantly applied throughout the cultivation, only 60% of the cells that expressed *CFP* and *RFP* from plasmids produced both proteins at detectable levels. If the presence of three plasmids is required in order for all proteins to be co-expressed in a cell, the proportion of cells that co-produces all three proteins drops dramatically. In the previous (**Chapter 3**) it was shown that only 6% of all analyzed yeast population are capable to co-produce three (CFP, RFP and YFP) proteins in the same cell [21].

All aforementioned advantages of the GA system are worthless if strain construction is time consuming. Fortunately, the procedure of GA was optimized to be as short as ten days (not including the time for DNA substrate construction and transformation). Moreover, due to smart design of the GA cassette no intensive screening efforts are required to identify correct clones as all strains where gene amplification was unsuccessful or incomplete are automatically selected against in the screening system depicted in the **Figure 4.5**. In fact, routinely three out of four plausible strains with amplified genes appear to be correct after checking with analytical PCR (no data).

To this end, the GA system was shown to be suitable for achieving high gene expression levels which is a prerequisite when constructing a microbial production host. In addition, the precise control of copy number and genomic location of amplified genes in combination with fast strain construction makes the GA system suitable for metabolic engineering of heterologous pathways in yeast. In fact, the next Chapter (**5**) of this thesis demonstrates the successful application of GA system for identification and elimination of metabolic bottlenecks in the *de novo* biosynthetic pathway for production of vanillin glucoside.

Conclusion

In summary, the gene amplification developed in this study was proven to be robust and easy to manipulate, and successful amplification of the reporter *CFP* and *RFP* genes in up to ten copies in haploid strains and in up to 20 in diploid strains was demonstrated. In some cases, protein production levels obtained in GA strains were found to be higher than the ones obtained in the strains expressing the same genes from an episomal multicopy plasmids. It was also shown that overexpression of two genes via the gene amplification system leads to a more homogeneous expression within the yeast population compared to plasmid based expression from 2μ plasmids. Moreover, the GA strains were shown to be able to sustain stable expression of amplified genes in long term cultivations. To this end, the platform developed in this work can be well-suited for the construction of genetically stable cell factories where single or multiple genes have to be overexpressed; in addition, the GA can be applied for metabolic engineering purposes where precise control of gene expression levels are needed.

Future perspectives

This chapter describes only a fraction of all the work being done for this particular project. An ongoing research also involves studies of two additional model systems for further validation and characterization of the GA method. Firstly, in order to precisely evaluate protein production in GA strains, the *lacZ* gene will be amplified to various copy numbers in the same way as it was described above for the fluorescent proteins. By using conventional methods β-galactosidase concentration can be easily determined and more importantly normalized to the cell dry-weight. The latter is especially important when comparing protein production in haploid and diploid cells since their size and volume are different. The β-galactosidase production in GA strains will also be compared to plasmid based systems. Next, a simple heterologous two component biosynthetic pathway for production of the polyketide 6-methylsalicylic acid (6-MSA) will be studied in GA strains. 6-MSA was previously produced in yeast by expressing the genes *npga* and *6-MSAS* from multi copy plasmids [22]. In this study, the two genes will be amplified in GA strains and 6-MSA production will be compared to the plasmid based platform. Moreover, 6-MSA pathway titration experiment will be done by amplifying independent parts of the pathway in order to investigate whether any (or which) of the two enzymes represent a bottleneck in the synthesis of 6-MSA.

Materials and Methods

Media, plasmids and yeast strains

Cultivation media: Genetic manipulations and cultivations of the yeast strains were done in a synthetic complete (SC) medium (Sherman et al. [23]) with minor modifications where the leucine concentration was doubled to 60 mg/L. All yeast transformants with gene integrations or plasmids were selected on SC media missing the appropriate auxotrophic marker (denoted as SC-aux, e.g. SC-trp if tryptophan is missing). In order to select against the *TRP1* or *URA3* markers, yeast strains were grown on SC medium plates supplemented with 500 mg/L 5-fluoroanthranilic acid (SC-5-FAA) or 740 mg/L 5-fluororotic acid (SC-5-FOA), respectively. For gene amplification purposes (induction of I-SceI plasmid) SC medium was modified by using 2% galactose instead of glucose as a carbon source and denoted as SCgal. Yeast Extract Peptone Dextrose (YPD) medium plates were used for storage and propagation of the constructed yeast strains as well as for sexual crossing purposes.

Plasmid construction: The I-SceI expression plasmid pWJ1320 was obtained from Michael Lisby [16]. Plasmids used in this study were assembled by USER™ cloning technique previously described in [24], [25]. The DNA fragments for cloning were amplified by PCR from the appropriate vector templates or genomic DNA of *S. cerevisiae* using PfuX7 polymerase developed by Nørholm et al. [26] with specially designed primer pairs (**SI Table S 4.1**). A detailed list of all plasmids used and constructed in this work is depicted in **SI Table S 4.2**.

Strain construction: The *Saccharomyces cerevisiae* strains CEN.PK110-4C and CEN.PK113-6B were used as starter strains for construction of all GA strains. For plasmid comparison experiments CEN.PK113-1C and CEN.PK111-61A were used. All CEN.PK strains were obtained from Peter Kötter [27]. The GA strains were constructed by sequential transformation with specifically designed GA cassettes targeting unique locations on three chromosomes X, XI and XII. Later, several rounds of crosses were performed to build strains containing up to ten GA cassettes (see **SI Figure S 4.1** and **Figure S 4.2**). For the proof of concept experiments, the strains with different copy numbers of *CFP* and *RFP* were constructed by applying the GA technology (see below for description); or by transforming WT CEN.PK strains with episomal multi-copy plasmids harboring the *CFP* or *RFP* genes. The GA strains expressing up to 20 copies of *CFP* and *RFP* were generated by mating opposite mating type haploids with different copy numbers of the reporter genes (see **SI Table S 4.4**), the diploids were constructed using the method described in [28].

Gene amplification protocol

For a schematic illustration of the gene amplification protocol see **Figure 4.5**.

Preparation of gene targeting fragments: Genes of interest (G1 and/or G2) are first assembled by USER cloning into the backbone of amplification plasmid pCSN. The final DNA construct containing the gene/s of interest, the selection marker TRP1 (flanked by the direct repeats) and the recombination sequences A and B are prepared by digesting the pCSN based vector with the relevant insert using NotI restriction endonuclease. Lastly, the relevant DNA fragment is purified from an agarose gel.

Transformation 1: A strain with a selected number of GA cassettes is transformed with the gene targeting substrate containing the gene(s) to be amplified. Transformants are selected and streak-purified on plates with SC-trp medium. Next, the TRP1 marker is recycled by purify streaking transformants on plates with SC-5-FAA medium.

Transformation 2: The resulting strain from **transformation 1** is transformed with the plasmid pWJ1320-TRP, which encodes the I-SceI endonuclease under the control of the galactose inducible *GAL1* promoter. The resulting transformants are selected on SC-trp medium. In this step it is important to adjust the concentration of pWJ1320-TRP plasmid in order to get the right amount of colonies on the plates (preferably 50 – 200 well separated colonies).

Induction: SC-trp plates with 50-200 single colonies are replicated onto induction media (SCgal-trp) containing plates. The plates are incubated at 30°C were for 2-4 days, the time of incubation can vary depending on the product properties of the amplified gene.

Selection: To identify the clones where gene amplification was successful, the SCgal-trp plates are replicated twice. Firstly, the SCgal-trp plate is replicated onto SC-5-FOA and incubated for 2-3 days at 30°C. Secondly, the SC-5FOA plate is replica plated onto YPD and SC-ura (lacking uracil) both plates are incubated overnight at 30°C. Colonies growing on YPD plates but not on SC-ura are plausible candidates with successfully amplified genes of interest. A reasonable number of colonies is streak-purified on plates with SC-5-FOA medium.

Verification: Finally, the successful amplification is validated by PCR on genomic DNA with the primers pairs (see **SI Table S 4.1**) targeting for the specific amplification sites.

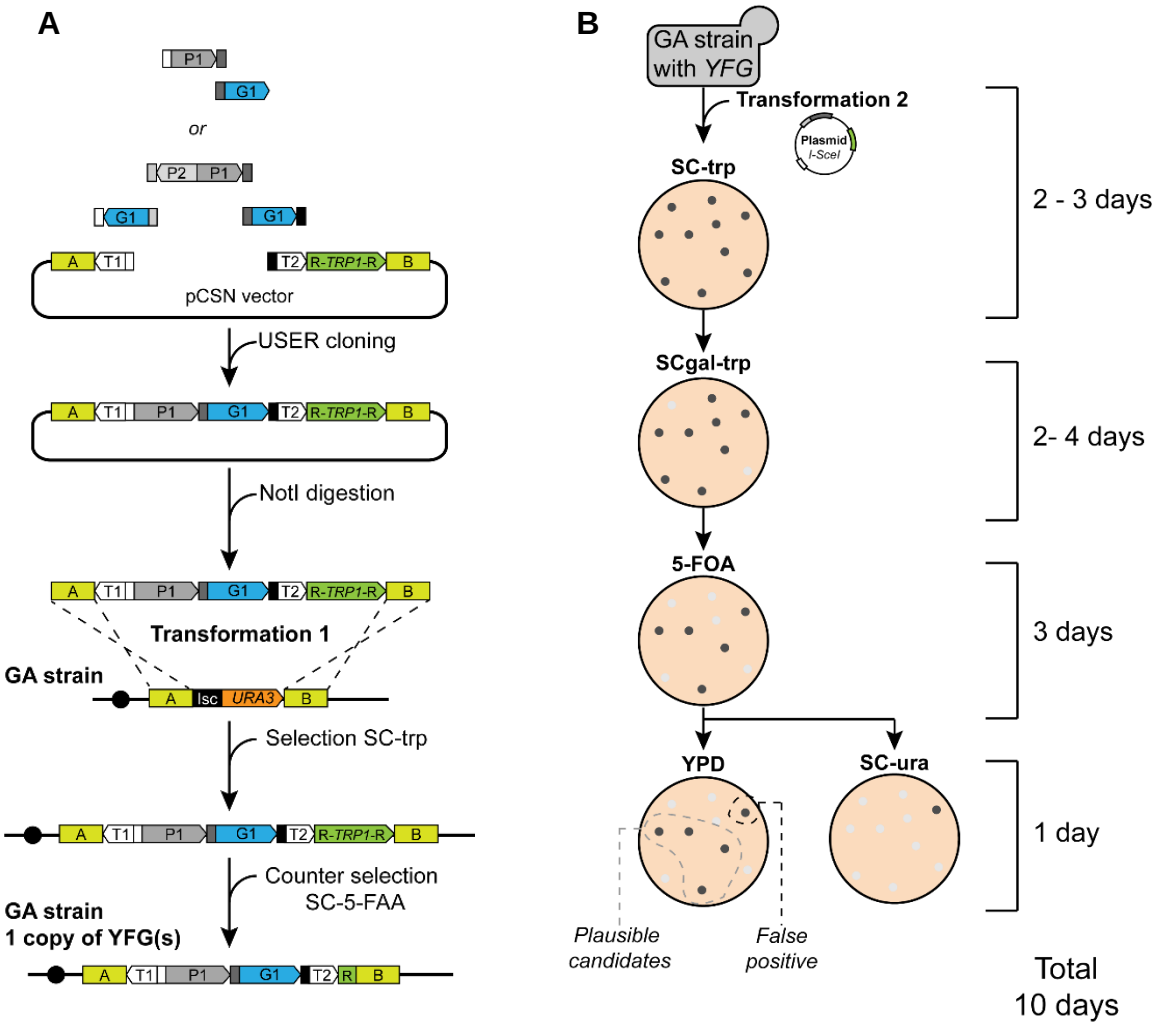


Figure 4.5.Generalized illustration of the gene amplification method. **A**) Assembly of the DNA fragment to be amplified and its transformation into GA strain with N-copies of amplification cassettes. **B**) Induction of the gene amplification and subsequent selection of the strains with successfully amplified genes. T1 – *ADH1* terminator, T2 – *TRP1* terminator, A and B – recombination sequences targeting GA cassette, P1, P2 – any promoter and G1, G2 – any gene of interest.

Flow cytometry analysis

Sample preparation: Pre-cultures of the plasmid containing and the GA strains were grown O/N in 3 mL SC medium (for plasmid SC-his-ura). 50 µL of pre-culture were used to inoculate 15 mL of fresh medium in 50 mL falcon tubes, where the cells were grown at 30°C with 150 rpm agitation at 30°C. After the cultures had reached mid-exponential phase they were harvested and fixed with paraformaldehyde according to the following protocol. 2 mL samples were taken, cooled on ice and subsequently centrifuged at 4°C, 3000 x g for 2 min. Supernatant was removed and pellet was resuspended in 200 µL of 2 % paraformaldehyde. The mix was incubated on ice for 1 hour and subsequently centrifuged at 4°C, 3000 x g for 2 min. Finally, the paraformaldehyde was removed

and pellet was resuspended in 200 µL PBS buffer. The fixed cells were stored at 4°C until FACS analysis.

Sample analysis: The cell samples were analyzed on a BD FACSAria equipped with three solid state diode lasers: air-cooled Coherent™ Sapphire™ solid-state diode laser (488 nm, 100 mW), air-cooled Coherent™ Yellow Green laser (561 nm, 100 mW), and an air-cooled Coherent™ Deep Blue laser (445 nm, 50 mW). The following filters were used: mCFP-A and PE-Cy5-A for analysis of emission from CFP and RFP, respectively. Signal compensation was performed according to manufacturer's protocol (BD FACSAria II User's Guide). For a control samples, strains without fluorescent proteins and two strains producing each fluorescent protein independently were used. Flow cytometry data sets were analyzed and interpreted by open source software (Flowing Software v2.5.0) developed by Perttu Terho.

Genetic stability assay

Triplicates of multi-copy plasmid strain and 10 copy strain (GAx10) expressing *CFP* and *RFP* were grown sequentially in 500 mL shake flasks containing 100 mL of SC (for plasmid strains SC-ura-his) medium. Cultures were incubated at 30°C with 150 rpm agitation, 2 mL samples for FACS analysis were collected once a day during mid-exponential growth phase and processed for storage using paraformaldehyde fixation method described above. Simultaneously, small volume of the each culture was inoculated to the new shake flask (starting OD_{600}=0.01) with the fresh media. The cycle was repeated for five days, all accumulated samples were analyzed in flow cytometer at the same day using the settings described above.

References

[1] G. Liti, D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. a James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman, D. Bensasson, M. J. T. O'Kelly, A. van Oudenaarden, D. B. H. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. a Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, and E. J. Louis, "Population genomics of domestic and wild yeasts.," *Nature*, vol. 458, no. 7236, pp. 337–41, Mar. 2009.

[2] T.-M. Lo, T. W. Suong, H. Ling, B. Chen, A. Kang, and M. W. Chang, "Microbial engineering strategies to improve cell viability for biochemical production.," *Biotechnol. Adv.*, pp. 1–12, Feb. 2013.

[3] A. H. Kachroo, M. Jayaram, and P. a Rowley, "Metabolic engineering without plasmids.," *Nat. Biotechnol.*, vol. 27, no. 8, pp. 729–31, Aug. 2009.

[4] P. M. Kilonzo, A. Margaritis, and M. a Bergougnou, "Plasmid stability and kinetics of continuous production of glucoamylase by recombinant Saccharomyces cerevisiae in an airlift bioreactor.," *J. Ind. Microbiol. Biotechnol.*, vol. 36, no. 9, pp. 1157–69, Sep. 2009.

[5] Z. Zhang, M. Moo-Young, and Y. Chisti, "Plasmid stability in recombinant Saccharomyces cerevisiae.," *Biotechnol. Adv.*, vol. 14, no. 4, pp. 401–35, Jan. 1996.

[6] P. Caunt, A. Impoolsup, and P. F. Greenfield, "Stability of recombinant plasmids in yeast," *J. Biotechnol.*, vol. 8, no. 3, pp. 173–192, Jul. 1988.

[7] T. S. Lopes, J. Klootwijk, A. E. Veenstra, P. C. Van Der Aarb, H. Van, H. A. Rauc, and R. J. Planta, "High-copy-number integration into the ribosomal DNA of Saccharomyces cerevisiae : a new vector," *Gene*, vol. 79, pp. 199–206, 1989.

[8] A. Sakai, Y. Shimizu, and F. Hishinuma, "Integration of heterologous genes into the chromosome of Saccharomyces cerevisiae using a delta sequence of yeast retrotransposon Ty," *Appl. Microbiol. …*, pp. 302–306, 1990.

[9] T. S. Lopes, I. J. de Wijs, S. I. Steenhauer, J. Verbakel, and R. J. Planta, "Factors affecting the mitotic stability of high-copy-number integration into the ribosomal DNA of Saccharomyces cerevisiae.," *Yeast*, vol. 12, no. 5, pp. 467–77, Apr. 1996.

[10] X. Wang, Z. Wang, and N. a Da Silva, "G418 Selection and stability of cloned genes integrated at chromosomal delta sequences of Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 49, no. 1, pp. 45–51, Jan. 1996.

[11] R. Parekh, M. Shaw, and K. Wittrup, "An integrating vector for tunable, high copy, stable integration into the dispersed Ty δ sites of Saccharomyces cerevisiae," *Biotechnol. Prog.*, 1996.

[12] K. E. J. Tyo, P. K. Ajikumar, and G. Stephanopoulos, "Stabilized gene duplication enables long-term selection-free heterologous pathway expression.," *Nat. Biotechnol.*, vol. 27, no. 8, pp. 760–5, Aug. 2009.

[13] J. D. Boeke, H. Xu, and G. R. Fink, "A general method for the chromosomal amplification of genes in yeast.," *Science*, vol. 239, no. 4837, pp. 280–2, Jan. 1988.

[14] F. W. Lee and N. a Da Silva, "Ty1-mediated integration of expression cassettes: host strain effects, stability, and product synthesis.," *Biotechnol. Prog.*, vol. 12, no. 4, pp. 548–54, 1996.

[15] M. D. Mikkelsen, L. D. Buron, B. Salomonsen, C. E. Olsen, B. G. Hansen, U. H. Mortensen, and B. A. Halkier, "Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform.," *Metab. Eng.*, vol. 14, no. 2, pp. 104–11, Mar. 2012.

[16] M. Lisby, U. H. Mortensen, and R. Rothstein, "Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre.," *Nat. Cell Biol.*, vol. 5, no. 6, pp. 572–7, Jun. 2003.

[17] B. G. Hansen, B. Salomonsen, M. T. Nielsen, J. B. Nielsen, N. B. Hansen, K. F. Nielsen, T. B. Regueira, J. Nielsen, K. R. Patil, and U. H. Mortensen, "Versatile enzyme expression and characterization system for Aspergillus nidulans, with the Penicillium brevicompactum polyketide synthase gene from the mycophenolic acid gene cluster as a test case.," *Appl. Environ. Microbiol.*, vol. 77, no. 9, pp. 3044–51, May 2011.

[18] J. Haber, "Partners and pathways: repairing a double-strand break," *Trends Genet.*, vol. 16, no. 6, pp. 259–264, 2000.

[19] F. Fang, K. Salmon, M. W. Y. Shen, K. A. Aeling, E. Ito, B. Irwin, U. P. C. Tran, G. W. Hatfield, N. A. Da Silva, and S. Sandmeyer, "A vector set for systematic metabolic engineering in Saccharomyces cerevisiae," *Yeast*, vol. 3, no. October 2010, pp. 123–136, 2011.

[20] A. S. Karim, K. a Curran, and H. S. Alper, "Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications.," *FEMS Yeast Res.*, vol. 13, no. 1, pp. 107–16, Feb. 2013.

[21] N. B. Jensen, T. Strucko, K. R. Kildegaard, F. David, J. Maury, U. H. Mortensen, J. Förster, J. B. Nielsen, and I. Borodina, "EasyClone: method for iterative chromosomal integration of multiple genes in Saccharomyces cerevisiae," *FEMS Yeast Res.*, pp. DOI:10.1111/1567–1364.12118, 2013.

[22] S. Wattanachaisaereekul, A. E. Lantz, M. L. Nielsen, O. S. Andresson, and J. Nielsen, "Optimization of heterologous production of the polyketide 6-MSA in Saccharomyces cerevisiae," *Biotechnol. Bioeng.*, vol. 97, no. 4, pp. 893–900, 2007.

[23] F. Sherman, G. R. Fink, and J. B. Hicks, *Laboratory course manual for methods in yeast genetics*. Cold Spring Harbor Laboratory, 1986.

[24] F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, and B. a Halkier, "USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.," *Nucleic Acids Res.*, vol. 35, no. 7, p. e55, Jan. 2007.

[25] H. H. Nour-Eldin, B. G. Hansen, M. H. H. Nørholm, J. K. Jensen, and B. a Halkier, "Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments.," *Nucleic Acids Res.*, vol. 34, no. 18, p. e122, Jan. 2006.

[26] M. H. H. Nørholm, "A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering.," *BMC Biotechnol.*, vol. 10, p. 21, Jan. 2010.

[27] K. Entian and P. Kötter, "Yeast genetic strain and plasmid collections," *Methods Microbiol.*, vol. 36, no. 06, pp. 629–666, 2007.

[28] D. a Treco and F. Winston, "Growth and manipulation of yeast.," *Curr. Protoc. Mol. Biol.*, vol. Chapter 13, no. April, p. Unit 13.2, Apr. 2008.

[29] R. J. D. R. Reid, M. Lisby, and R. Rothstein, "Cloning-free genome alterations in saccharomyces cereuisiae using adaptamer-mediated PCR," *Methods Enzymol.*, vol. 350, no. 1983, pp. 258–277, 2002.

Supplementary Information

Contents

List of Supplementary Figures

List of Supplementary Tables

Supplementary methods

Gene amplification strains were constructed as shown in **Figure S 4.1** and **Figure S 4.2**.

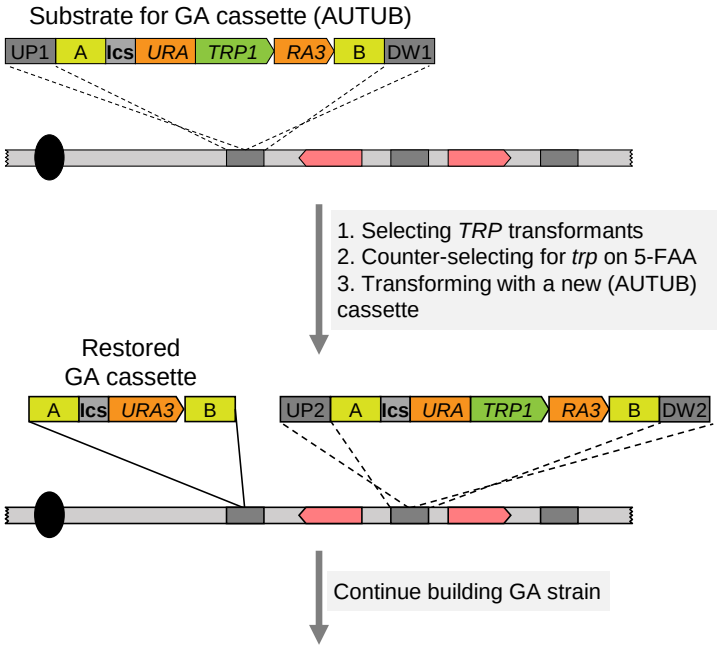


Figure S 4.1. Simplified representation of gene amplification (GA) cassette introductions into yeast genome. Targeting substrates named "AUTUB" were generated by USER cloning, all required fragments (A, B – 500 bp sequences from *Neurospora crassa*, Ics – homing endonuclease (IsceI) recognition sequence, *URA* and *RA3* – 3' and 5' truncated sequences (containing overlapping regions) of *URA3* gene from *Kluyveromyces lactis*, respectively, and selection marker *TRP1* were assembled into ten backbone vectors with specific targeting regions (UP and DW) into *S. cerevisiae* genome [15]. The GA cassettes are integrated by sequential transformations with AUTUB fragments and selecting the resulting strains on SC-trp medium. Subsequently, *TRP1* marker is counter selected on SC-5-FAA medium where resulting strain harbors GA cassette with the functional *URA3* gene. Once, the first GA cassette is integrated, the following transformation involves the integration of next AUTUB substrate. Cycle is repeated until required number of the GA cassettes is achieved. Red arrow shaped boxes represent essential genetic elements separating the integration sites [15].

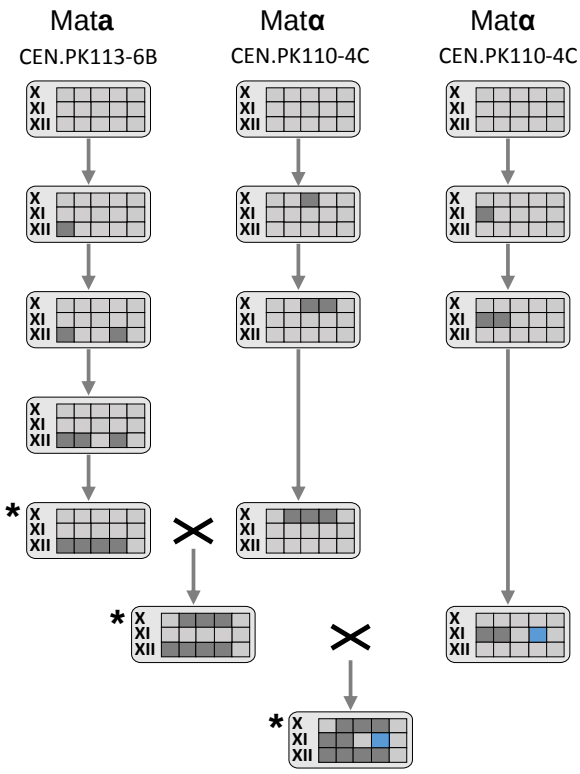


Figure S 4.2. Construction of the GA strain collection. The GA cassettes were integrated as previously described in **Figure S 4.1**. Three traits of strains were built simultaneously where evolution of amplification cassettes was done on three separate chromosomes X, XI and XII represented by rows of light grey boxes. The dark grey box depicts location where the GA cassette is integrated (the number of the column corresponds to the number of the integration site describe by Mikkelsen [15]). The blue box represents integration site characterized in this study. Next, opposite mating type strains undergone the series of crosses which resulted in construction of strains with up to ten GA cassettes. (*) represent a collection of strains with different mating types and selection markers. For the list of the strains refer to the **Table S 4.3**.

Table S 4.1. Primers used in this study.

Name	Sequence
A-fwd-U1A-adap	CGTGCGAUAGGTGTAAAAGTAGGGAGCG
A-rv	TCACTGTCCGCTTGCAGAGT
TADH1-fwd-A-adap	ACTCTGCAAGCGGACAGTGAAGGGAAGAAAGCGAAAGGAG
TADH1-Rv-U2-adap	ATGCACGCGAUCGCACGCATTCGGGTCAAATCGTTGGTAGATACG
C-fwd-U2-adap	ATCGCGTGCAUTCCTATCTTATGCCTTCATTTC
Ny-C-rv-B-adap	CAGGAGTAGCGTACATAGGACAAAGCCCATTTCAATGAGC
B-fwd	TCCTATGTACGCTACTCCTG
B-rv-U1B-adap	CACGCGAUGAGGAGAGTGGATGGATAGTC
Trp-L1-fwd	AGTCAGTUTGGCTTAACTATGCGGCATC
Trp-U3-rv	CACGCGAUCCTTTGACGTTGGAGTCCAC
Trp-U3-AsiSI-repeat-fwd	CGTGCGAUGAATGCGTGCGATCGCGTGCATTCGTTTGTGCACTTGCCTATGC
Trp-L1-repeat-rv	AACTGACUCCTTTGACGTTGGAGTCCAC

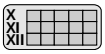
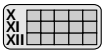
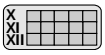
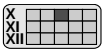
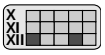
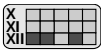
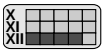
Name	Sequence
A-fwd-AsiSI/Bsm-adap	cgtgcga**U**AGGTGTAAAAGTAGGGAGCG
A-fwd-AsiSI/Bsm-adap	cgtgcga**U**AGGTGTAAAAGTAGGGAGCG
3-int-U1-adap	acgaatgc**U**GAGCAATGAACCCAATAACGAAATC
Trp1-up-fwd-U1-adap	agcattcg**U**TGGCTTAACTATGCGGCATC
Trp1-down-Rv-U2-adap	agtagcat**U**CCTTTGACGTTGGAGTCCAC
5-int-U2-adap	aatgctac**U**CTTGACGTTCGTTCGACTGATGAGC
B-rv-AsiSI/Bsm-cass-adap	cacgcga**U**GAGGAGAGTGGATGGATAGTC
A-fwd-U3-adap	agctacg**U**AGGTGTAAAAGTAGGGAGCG
B-Rv-U4-adap	agcaagg**U**GAGGAGAGTGGATGGATAGTC
X-2-UP-rv-U3-adap	acgtagc**U**ACGTGACCACTTCGAGAGCA
X-2-DW-fwd-U4-adap	accttgc**U**CCTGCATAATCGGCCTCACA
X-3-UP-rv-U3-adap	acgtagc**U**AGTCTCGTATGTCGGCTCTC
X-3-DW-fwd-U4-adap	accttgc**U**TGTGTCCGCGTTTCTAAGGC
X-4-UP-rv-U3-adap	acgtagc**U**CTGCTCTTGAATGGCGACAG
X-4-DW-fwd-U4-adap	accttgc**U**AACAGGCATGGGAAGATTCGC
XI-1-UP-rv-U3-adap	acgtagc**U**GGCTGCAAAGAATCCTCCGAG
XI-1-DW-fwd-U4-adap	accttgc**U**CTTCCACGGAATACCAAGCCC
XI-2-UP-rv-U3-adap	acgtagc**U**TGTGGAAGTTCATGGCAAACG
XI-2-DW-fwd-U4-adap	accttgc**U**CACACGACTAGCGCTTTCAG
XI-3-UP-rv-U3-adap	acgtagc**U**AATCAGACGCACGCTTGGCG
XI-3-DW-fwd-U4-adap	accttgc**U**TACGTGGATTGAGCCAGCAATAC
XII-1-UP-rv-U3-adap	acgtagc**U**GAAAGAACCGAACCGATGCCA
XII-1-DW-fwd-U4-adap	accttgc**U**CTTCCCGTGAATCAACTGCAC
XII-2-UP-rv-U3-adap	acgtagc**U**CATAACGCGTTACACGGAAGG
XII-2-DW-fwd-U4-adap	accttgc**U**CTACTATCGGCGACTCTCTC
XII-3-UP-rv-U3-adap	acgtagc**U**GTAATGCGAATGAGCAGGTAC
XII-3-DW-fwd-U4-adap	accttgc**U**GGAAGTTTTGCAGATGAAGTGC
XII-4-UP-rv-U3-adap	acgtagc**U**CTATCGCTGAACAGGAACTTAAG
XII-4-DW-fwd-U4-adap	accttgc**U**CCAACTTTGTACTATTCCTTCCC
XII-5-UP-rv-U3-adap	acgtagc**U**GCTGATGTGACACTGTGACAA
XII-5-DW-fwd-U4-adap	accttgc**U**GCAACTCAGAAGTTTGACAGCAAG
Verification primers	
XI-4A_UF	GGGTTTAA**U**GTGAATACTGCACAAAATGAATAC
XI-4A_UR	GGACTTAA**U**CTTGAATCTTTTAAGCAGCCA
XI-4A_GA_UR	Acgtagc**U**CTTGAATCTTTTAAGCAGCCA
XI-4_DF	GGCATTAA**U**ATCAAGAGAAACATTACAAGGC
XI-4_GA_DF	accttgc**U**ATCAAGAGAAACATTACAAGGC
XI-4A_DR	GGTCTTAAUCTGTGGCAAGATTGGCAAG
XI-4A_out_DR	CAGCAACCAGAATATCTATGATG
X-2-up-out-sq	TGCGACAGAAGAAAGGGAAG
X-3-up-out-sq	TGACGAATCGTTAGGCACAG
X-4-up-out-sq	CTCACAAAGGGACGAATCCT
Ny-XI-1-up-sq	CTTAATGGGTAGTGCTTGACACG
XI-2-up-out-sq	CAATATCAGTGTGGTGAAC
XI-4A_out_UF	AAGCCCTATTATTGCTGACTTG
XII-1-up-out-sq	CTGGCAAGAGAACCACCAAT
XII-2-up-out-sq	CGAAGAAGGCCTCCAATTC

Name	Sequence
XII-3-up-out-sq	TGGGCAGCCTTGAGTAAATC
XII-4-up-out-sq	GAACTGACGTCGAAGGCTCT
C1_TADH1_F	CTTGAGTAACTCTTTCCTGTAGGTC

Table S 4.2. List of the plasmids used and created in this study.

Name	Description	Reference
Template plasmids		
pX-2	*Backbone for targeting to X2*	[15]
pX-3	*Backbone for targeting to X3*	[15]
pX-4	*Backbone for targeting to X4*	[15]
pXI-1	*Backbone for targeting to XI1*	[15]
pXI-2	*Backbone for targeting to XI2*	[15]
pXI-4A	*Backbone for targeting to XI4A*	This study
pXII-1	*Backbone for targeting to XII1*	[15]
pXII-2	*Backbone for targeting to XII2*	[15]
pXII-3	*Backbone for targeting to XII3*	[15]
pXII-4	*Backbone for targeting to XII4*	[15]
pWJ1042	*TRP1*	[29]
pWJ1163	*CFP*	[29]
pWJ1350	*RFP*	[16]
Plasmids for GA cassette introduction		
pX2-AUTUB	*Targets GA cassette to X2*	This study
pX3-AUTUB	*Targets GA cassette to X3*	This study
pX4-AUTUB	*Targets GA cassette to X4*	This study
pXI1-AUTUB	*Targets GA cassette to XI1*	This study
pXI2-AUTUB	*Targets GA cassette to XI2*	This study
pXI4A-AUTUB	*Targets GA cassette to XI4*	This study
pXII1-AUTUB	*Targets GA cassette to XII1*	This study
pXII2-AUTUB	*Targets GA cassette to XII2*	This study
pXII3-AUTUB	*Targets GA cassette to XII3*	This study
pXII4-AUTUB	*Targets GA cassette to XII4*	This study
Plasmids targeting GA cassette		
pCSN	AsiSI/Nb.BsmI cassette	This study
pCSN-CFP-RFP	p*TPI*::*CFP,* p*ACT*::*RFP*	This study
Muticopy episomal plasmids		
pESC-HIS	2µ, *HIS3*	Agilent
pESC-URA	2µ, *URA3*	Agilent
pESC-his-RFP	p*ACT*::*RFP*, 2µ	This study
pESC-ura-CFP	p*TPI*::*CFP*, 2µ	This study
IsceI inductive plasmids		
pWJ1320	p*GAL1*:*IsceI*	[16]
pWJ1320-TRP1	p*GAL1*:*IsceI*	This study
pWJ1320-HIS3	p*GAL1*:*IsceI*	This study
pWJ1320-LEU2	p*GAL1*:*IsceI*	This study
pWJ1320-KanMX	p*GAL1*:*IsceI*	This study
pWJ1320-HghMx	p*GAL1*:*IsceI*	This study

Table S 4.3. List of the strains used and constructed in this work. All strains are CEN.PK based, the genotype of all strains is MAL2-8C SUC2 + (definition in the table). The colored square in the grey 3X5 array represents the location of the amplification cassette or amplified genes of interest, each column of the array corresponds to the number of the integration site described earlier by Mikkelsen et al. (2012).

NAME	Mat**a**	Mat**α**	*his3*	*leu2*	*trp1*	*ura3*	Copy No.	Cassette location	Description or Reference
Starting strains									
CEN.PK110-4C		•	•		•	•	—		Peter Kötter [27]
CEN.PK113-6B	•			•	•	•	—		Peter Kötter [27]
CEN.PK113-1C	•		•		•	•	—		Peter Kötter [27]
CEN.PK113-7A	•		•				—		Peter Kötter [27]
CEN.PK113-7B		•		•	•	•	—		Peter Kötter [27]
CEN.PK111-61A		•	•	•		•	—		Peter Kötter [27]
Gene amplification strains									
GA-X-3		•	•		•		1		Constr. by transformation
GA-XI-1		•	•		•		1		Constr. by transformation
GA-XII-1	•			•	•		1		Constr. by transformation
GA-X-3+4		•	•		•		2		Constr. by transformation
GA-XI-1+2		•	•		•		2		Constr. by transformation
GA-XII-1+4	•			•	•		2		Constr. by transformation
CSN-9C		•	•		•		2		Constructed by cross
GA-X-3+4+2		•	•		•		3		Constr. by transformation
GA-XI-1+2+4a		•	•		•		3		Constr. by transformation
GA-XII-1+4+2	•			•	•		3		Constr. by transformation
GA-XII-1+4+2+3	•			•	•		4		Constr. by transformation
GAX4-2B; 3A		•		•	•		4		Constructed by cross
GAX4-2D	•		•		•		4		Constructed by cross
GAX4-3D		•			•		4		Constructed by cross
GAX7-5C; 34D; 43C		•		•	•		7		Constructed by cross
GAX7-23A		•	•		•		7		Constructed by cross
GAX7-12B; 34C	•		•		•		7		Constructed by cross

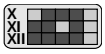
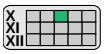
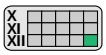
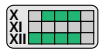
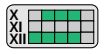
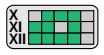
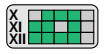
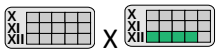
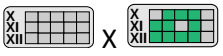
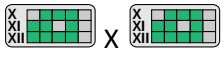
NAME	Mata	Matα	*his3*	*leu2*	*trp1*	*ura3*	Copy No.	Cassette location	Description or Reference
GAX7-5D	•		•	•	•		7		Constructed by cross
GAX10-1C; 4A; 9A	•		•		•		10		Constructed by cross
GAX10-2B; 4B; 9D		•		•	•		10		Constructed by cross
GAX10-7A		•	•		•		10		Constructed by cross
GAX10-4D	•		•	•	•		10		Constructed by cross
A proof of concept strains with amplified CFP and RFP genes									
H0+FP (113-1C)	•		•		•	•	—		Negative control
H1+CFP	•		•		•	•	1		One color control CFP
H1+RFP	•		•		•	•	1		One color control RFP
GAX7-34C+FP unamp.	•		•		•	•	1		Not amplified 7 copy strain
CSN-9C+FP; 1A; 2A	•		•		•	•	2		A proof-of-concept strain
GAX4-1A-FP	•		•		•	•	4		A proof-of-concept strain
GAX7-5C+FP; 1D; 2C		•		•	•	•	7		A proof-of-concept strain
GAX7-34C+FP; 1A; 4A	•		•		•	•	7		A proof-of-concept strain
GAX10-1C+FP;1A; 2B	•		•		•	•	10		A proof-of-concept strain
GAX10-2B+FP; 2; 5		•		•	•	•	10		A proof-of-concept strain
H2u-FP (113-1C)	•				•		?	Plasmids	Multi copy plasmid expression (pESC-his-RFP, pESC-ura-CFP)

Table S 4.4. The diploid strain construction containing various copy number of amplified *CFP* and *RFP* genes.

Name	Mat a strain	Mat alpha strain	Copies	Comments
D0-FP	CEN.PK113-7A	CenPK113-7B	0	Non fluorescence control
D4-FP	GAx4-FP-1A	CEN.PK113-7B	4	
D10-FP	GAX10-1C+FP	CEN.PK113-7B	10	
D14-FP	GAX7-34C+FP	GAX7-5C+FP	14	
D20-FP-A	GAX10-1C+FP, 1A	GAX10-2B+FP, 2	20	
D20-FP-B	GAX10-1C+FP, 2B	GAX10-2B+FP, 5	20	
D2u-FP	CEN.PK113-1C	CenPK111-61A	2µ	Harbors two plasmids pESC-his-RFP and pESC-ura-CFP

Chapter 5

Balancing vanillin production in yeast

Title[4]

Balancing the heterologous biosynthetic pathway for improved production of Vanillin-β-glucoside in *Saccharomyces cerevisiae*

Abstract

It was previously demonstrated that one of the most important aroma compounds – vanillin can be produced in baker's yeast in its glycosylated form – vanillin β-D-glucoside. Nevertheless, the amount of vanillin-β-glucoside obtained by the engineered cell factory was insufficient for commercial production. Metabolic engineering efforts to enhance the production were limited by the inherent capacity of the heterologous pathway to accommodate efficient conversion towards vanillin-β-glucoside. To identify the rate limiting steps of the vanillin-β-glucoside biosynthesis, we have systematically increased the dosage of each gene of the integrated pathway. Based on the experimental observation, subsequent overexpression of the gene combinations was performed. The most productive strain, combining the amplification of last two steps of the integrated biosynthetic pathway, conferred an approximately 6-fold increase in vanillin-β-glucoside levels. Moreover the accumulation of several intermediates was substantially decreased. Lastly, the data obtained in this study will enable further development of the improved vanillin-β-glucoside production in *S. cerevisiae.*

***Keywords**:* Yeast; Gene Amplification; Vanillin-β-glucoside; Pathway optimization

[4] Manuscript in preparation

Introduction

Vanillin is one of the most popular aromatic flavor and fragrance compounds, primarily used in food and perfumes and to some extent also for animal feed, tobacco and as precursor in the pharmaceutical industry. Originally, vanillin was extracted from the cured pods of the vanilla orchid*, Vanilla planifolia*, however, nowadays only about 0.25 % of the global annual market of approximately 16000 tons comes from the vanilla orchid. The remaining portion is mainly produced by chemical synthesis from lignin or fossil hydrocarbons such as guaiacol [1], [2]. As fossil based natural resources are decreasing rapidly and the need for "natural" vanillin is increasing, biosustainable and more environmentally friendly biotechnological routes were implemented. Those are mainly based on bioconversion of plant secondary metabolites (e.g. eugenol and ferulic acid) by various bacteria and fungi, and are summarized in reviews [3] and [4]. As an attractive alternative, Hansen and co-workers have assembled the *de novo* metabolic pathway for vanillin-β-glucoside (VG) production in yeast using glucose as a substrate [1]. VG is less toxic and more soluble than vanillin and is the form which naturally accumulates in the vanilla pod [2]. The engineered pathway involves the incorporation of five heterologous genes from several different and evolutionary distant organisms (see **Figure 5.1**). However, the amount of VG produced in the *S. cerevisiae* strain was insufficient for commercial production. In addition to that, subsequent studies on the VG biosynthetic pathway revealed significant accumulation of several intermediates, which, in fact, decreased the possible final yields of VG dramatically [5], [6]. Moreover, some intermediates, especially aromatic aldehydes, produced by the unbalanced heterologous pathway are toxic and can most likely compromise cell growth [1], [7].

Metabolic engineering has shown to be a successful tool in development and improvement of microbial cell factories for production of heterologous secondary metabolites [8]–[10]. Several approaches involving global and local metabolic engineering strategies have been considered in order to increase the VG production in *S. cerevisiae.* In the first case, the yeast central metabolism was reengineered to boost the supply of cofactors (in particular, ATP and NADPH) used by enzymes of the VG pathway, which led to fivefold improvement in production [5]. Subsequently, the *de novo* VG pathway was engineered by integrating an additional copy of either ACAR or HsOMT genes. While overexpressed ACAR had no effect on the VG metabolic profile, the additional copy of HsOMT resulted in slightly improved VG production by 30 % [6]. However, the problem of the accumulation of intermediates remained, indicating the presence of unsolved metabolic bottlenecks in the VG biosynthesis. Therefore, further improvements of the VG production in a yeast cell factory had to be considered.

The aim of this work was to systematically overexpress each VG pathway gene and several gene pair combinations in order to pinpoint metabolic bottlenecks in the biosynthesis of the VG. To achieve stable and controlled overexpression of the gene/s of interest, we have implemented a recently developed gene amplification (GA) system described in **Chapter 4**. Here, overexpression was assured by eight copies of the chromosomally integrated gene/s of interest. In addition, data obtained in this experiment provided a better understanding of the VG pathway dynamics which may give clues for further development of a yeast cell factory for VG production. Nonetheless, in this paper we have also demonstrated the first successful application of the GA for overexpression and manipulation of a complex heterologous pathway in yeast.

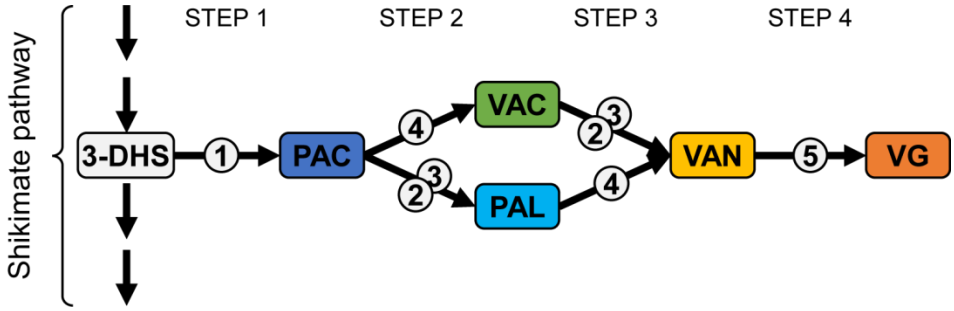


Figure 5.1. Simplified representation of the vanillin-β-glucoside (VG) production in *S. cerevisiae*. 3-DHS – 3-dehydroshikimic acid, PAC – protocatechuic acid, PAL – protocatechuic aldehyde, VAC – vanillic acid, VAN – vanillin, VG – vanillin-β-glucoside. Black thick arrows with numbered circles represent enzymatic reactions by heterologous enzymes: 1) 3DSD – 3-dehydroshikimate dehydratase (*Podospora anserina*), 2) ACAR – aromatic carboxylic acid reductase (*Neurospora sp.*), 3) PPTase (EntD) – phosphopantetheine transferase (*Escherichia coli*), 4) HsOMT – O-methyltransferase (*Homo sapiens*) and 5) UGT (UGT72E2) – UDP-glycosyltransferase (*Arabidopsis thaliana*).

Materials and Methods

Molecular cloning procedures

All molecular cloning procedures were done by applying the uracil-specific excision reagent (USER™) as previously described in [11], [12]. The genes constituting the *de novo* VG pathway and the promoters were amplified by PCR from the appropriate vector templates using PfuX7 polymerase developed by Nørholm et al. [13]. The pathway genes and promoters were assembled into the vector pCSN specially designed for the gene amplification system (GA) [**Chapter 4**]. A number of plasmids carrying single ORFs or their combinations were generated (see Table 5.1). Finally, all plasmids were validated by DNA sequencing (StarSEQ® GmbH, Germany).The full list of the plasmids used and constructed in this study is shown in the **Table 5.1**. The schematic plasmid cloning procedure is explained in **SI Figure S 5.3**.

Table 5.1. List of plasmids used in this study.

Name	Description	Reference
pCSN-01	For amplification of 3DSD	This study
pCSN-20	For amplification of ACAR	This study
pCSN-03	For amplification of EntD	This study
pCSN-04	For amplification of HsOMT	This study
pCSN-50	For amplification of UGT72E2	This study
pCSN-23	For amplification of ACAR and EntD	This study
pCSN-24	For amplification of ACAR and HsOMT	This study
pCSN-25	For amplification of ACAR and UGT72E2	This study
pCSN-54	For amplification of UGT72E2 and HsOMT	This study
pCSN	Backbone vector for amplification purposes	Chapter 4
pWJ1320-TRP	Inductive plasmid with Isce-I endonuclease	Chapter 4
pSP-G2	Template for bidirectional pPGK1/pTEF1 promoter	Partow et al. [14]
pJH500	Template for 3DSD gene	Hansen et al. [1]
pJH674	Template for ACAR gene	Hansen et al. [1]
pJH589	Template for EntD gene	Hansen et al. [1]
pJH543	Template for HsOMT gene	Hansen et al. [1]
pJH665	Template for UGT72E2 gene	Hansen et al. [1]
pUG6	Template for KanMX	Güldener et al [15]
pSH47	Plasmid containing Cre recombinase	Güldener et al [15]

Table 5.2. List of primers used in this study. All sequences are presented in 5' to 3' direction, standard capital letter are gene specific sequences, standard underline letters represent USER specific tails, *italics* underlined letters show targeting sequences for the appropriate gene deletions, and standard bold letters represents translational enhancer sequence [16], [17].

Name	Sequence
PGK_R-	ACCCGTTGAUGCCGCTTGTTTTATATTTGTTG
TEF_F+	CGTGCGAUGCCGCACACACCATAGCTTC
TEF_R+	ACGTATCGCUGTGAGTCGTATTACGGATCCTTG
DSD_F+	AGCGATACGU**AAAA**ATGCCTTCCAAACTCGCC
DSD_R+	CACGCGAUTTACAAAGCCGCTGACAGC
ACAR_F-	ATCAACGGGU**AAAA**ATGGCTGTTGATTCACCAGATG
ACAR_R-	CGTGCGAUCTTATAACAATTGTAACAATTCCAAATC
OMT_F+	AGCGATACGU**AAAA**ATGGGTGACACTAAGGAGCAA
OMT_R+	CACGCGAUCTTATGGACCAGCTTCAGAACC
PPT1_F+	AGCGATACGU**AAAA**ATGGTCGATATGAAAACTACGC
PPT1_R+	CACGCGAUTTAATCGTGTTGGCACAGC
UGT1_F-	ATCAACGGGU**AAAA**ATGCATATCACAAAACCACACG
UGT1_R-	CGTGCGAUACTAGGCACCACGTGACAAGTC
BGL1_del_F	*ATTTTTTGTTTACTTTCTTTTTCTAGTTAATTACCAACTAAA*CTTCGTACGCTGCAGGTC
BGL1_del_R	*CATTAGAAAATTCAGCTAAAATGAGCGGACTGAGGGCGA*CTAGTGGATCTGATATCACCTA
ADH6_del_F	*GAGGAAGAAATTCAACACAACAACAAGAAAAGCCAAAATC*CTTCGTACGCTGCAGGTC
ADH6_del_R	*GTTAAAAAGAAAGGAGCTACATTTATCAAGAGCTTGACAAC*TAGTGGATCTGATATCACCTA

Strain construction

The genotype and source of the strains used in this study is given **Table 5.3**. The yeast constructs used in this study were generated by three different techniques; first, by using a high efficiency genetic transformation method described by Gietz et al. [18],second, implementing a novel GA technique described by in **Chapter 4**, and third, by sexual mating following the protocol explained in [19].

Prior to construction of the strains with increased VG pathway gene copy number, several additional yeast strains had to be developed. The strain TS086 with double gene deletion (*Δadh6* and *Δbgl1*) was generated using a method described by Güldener et al [15]. For this reason, the strain CEN.PK110-7C was consecutively transformed with PCR fragments carrying the *loxP*-KanMX-*loxP* cassette amplified form plasmid pUG6 with the primers containing 40 nt long targeting sequences. The transformants were selected on YPD medium supplemented with G418 (sigma). In between the transformations, the KanMX marker was removed by expressing the Cre recombinase from the vector pSH47 [15].

Haploid strains (TS152 and TS155) containing the full VG biosynthetic pathway were obtained in a previous study by sexual cross of two CEN.PK110-16D and C-VG-aux strains, followed by sporulation and tetrad dissection (**Chapter 6**).

In this study the GA strain GAX7-34C (with integration capacity for seven copies) was used for overexpression purposes. In order for the GA to be compatible with the *de novo* VG pathway, a double gene deletion (*Δadh6* and *Δbgl1*) had to be introduced, which was done by crossing the strain GAX7-34C with TS086. The resulted diploid was then sporulated and after tetrad dissection the strain TS160 was selected for further genetic manipulations.

The haploid strains containing seven copies of the integrated gene of interest (GOI) were constructed in an identical manner for all combinations used in this study. The TS160 strain was transformed with a linear fragment obtained by digesting pCSN-## with *Not*I endonuclease (Fermentas-Thermo Fischer Scientific). Resulting transformants were selected and streak purified on SC-trp plates. Next, the TRP1 marker was eliminated by counter-selective plating on SC-5-FAA medium. Afterwards, these strains were transformed with the plasmid pWJ1320-TRP and subsequently selected on SC-trp plates. In order to induce the gene amplification process, strains with the pWJ1320-TRP plasmid were replica-plated on SCgal-trp. Finally, strains carrying seven copies of the GOI were selected by several consecutive replica-plating events using different selective media (the detailed gene amplification method is described in **Chapter 4**).

To analyze the effect of increased dosage of the VG pathway genes, diploid strains were constructed. This was done by mating haploid strains with the amplified GOIs with the haploid strain with the full VG biosynthetic pathway (**Table 5.4**). Diploid cells were selected and streak purified on SC-his-ura plates.

Finally, all gene targeting and gene amplification events were validated by diagnostic PCR using specially designed primer pairs (**SI Table S 5.2**)

Table 5.3. List of the haploid yeast strains used in this study.

Name	Genotype	Reference
CEN.PK110-7C	*MATα MAL2-8C SUC2 trp1-289 ura3-52*	Peter Kötter5
CEN.PK110-16D	*MATα MAL2-8C SUC2 trp1-289*	Peter Kötter1
GAX7-34C	*MATa MAL2-8C SUC2 his3Δ1 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-Ics-URA3-B))*	Chapter 4
C-VG-aux	*MATa SUC2 gal2 mal mel ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	Chapter 6
TS152	*MATa MAL2-8C SUC2 ura3-52 his3Δ XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	Chapter 6
TS155	*MATα MAL2-8C SUC2 ura3-52 his3Δ XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	Chapter 6
TS086	*MATα MAL2-8C SUC2 trp1-289 ura3-52 Δbgl1::loxP Δadh6::KanMX*	This study
TS160	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-Ics-URA3-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS192,TS193	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pTEF1-3DSD-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS199, TS201	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-ACAR-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS188, TS189	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pTEF1-PPT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS182, TS183	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pTEF1-HsOMT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS185, TS186	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-UGT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS186, TS187	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-ACAR,pTEF1-PPT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS250, TS251	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-ACAR,pTEF1-HsOMT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS252, TS253	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-ACAR,pTEF1-UGT-B)) Δbgl1::loxP Δadh6::KanMX*	This study
TS238, TS239	*MATα MAL2-8C SUC2 trp1-289 ura3-52 (X-2; X-3; X-4; XII-1; XII-2; XII-3; XII-4::(A-pPGK1-UGT,pTEF1-HsOMT-B)) Δbgl1::loxP Δadh6::KanMX*	This study

[5] Institut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

Table 5.4. Construction of diploid strains with amplified VG pathway genes.

Diploid*	Description	Haploid 1	Haploid 2
D-VG-00	Reference strain, one copy of VG pathway	TS086	TS152
D-VG-01	VG pathway with o/e 3DSD	TS192, TS193	TS152
D-VG-20	VG pathway with o/e ACAR	TS199, TS200	TS152
D-VG-03	VG pathway with o/e EntD	TS188, TS189	TS152
D-VG-04	VG pathway with o/e HsOMT	TS182, TS183	TS152
D-VG-50	VG pathway with o/e UGT72E2	TS185, TS186	TS152
D-VG-23	VG pathway with o/e ACAR and EntD	TS186, TS187	TS152
D-VG-24	VG pathway with o/e ACAR and HsOMT	TS250, TS251	TS152
D-VG-25	VG pathway with o/e ACAR and UGT72E2	TS252, TS253	TS152
D-VG-54	VG pathway with o/e UGT72E2 and HsOMT	TS238, TS239	TS152
D-VG-VG	Two copies of VG pathway	TS155	TS152

***** – represents two diploid strains as a results of two variants of haploid 1, which were used as a biological replicas in this experiment.

Media

To propagate *Escherichia coli* DH5α strains harboring the cloned plasmids, lysogeny broth (LB) [20] supplemented with 100 mg/L of ampicillin (Sigma) was used.

For genetic manipulations of yeast all media were prepared as previously described by Sherman et al. [21], with minor modifications of the synthetic complete (SC) medium where the leucine concentration was doubled to 60mg/L. All yeast transformants with gene integrations or plasmids were selected on synthetic complete media missing tryptophan (SC-trp). In order to remove the TRP1 marker, yeast strains were grown on SC medium plates supplemented with 500 mg/L 5-fluoroanthranilic acid (5-FAA).

For gene amplification purposes SC-trp medium was modified by using 2% galactose instead of glucose as a carbon source (SCgal-trp). Strains with amplified VG pathway genes were selected on SC medium containing 30 mg/L uracil and 740mg/L 5-fluoroorotic acid (5-FOA) (Sigma), while SC medium lacking uracil SC-ura was used as a negative control.

Yeast Extract Peptone Dextrose (YPD) medium plates were used for storage and propagation of the constructed yeast strains as well as for sexual crossing purposes. The medium composition was as follows: 10 g/L of yeast extract, 20 g/L of peptone and 20 g/L of glucose and 20 g/L of agar. For selection and propagation of engineered diploid strains with different combinations of amplified the VG pathway genes synthetic medium lacking histidine and uracil (SC-his-ura) was used.

For shake flask experiments a defined minimal medium previously described by Verduyn et al. [22] with 20 g/L glucose as a carbon source was used. The medium is composed of: 7.5 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.75 g/L Mg_2SO_4, 1.5 mL/L trace metal solution, 1.5 mL/L vitamins solution, 0.05 mL/L antifoam 204 (Sigma- Aldrich A-8311). The trace metal solution contains 3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 4.5 g/L $CaCl_2 \cdot 6H_2O$, 0.84 g/L $MnCl_2 \cdot 2H2O$, 0.3 g/L $CoCl_2 \cdot 6H_2O$, 0.3 g/L $CuSO_4 \cdot 5H_2O$, 0.4 g/L $NaMoO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 0.1 g/L KI and 15 g/L $Na_2EDTA \cdot 2H_2O$. The vitamin solution includes 50 mg/L d-biotin, 200 mg/L para-amino benzoic acid, 1.0 g/L nicotinic acid, 1.0 g/L Ca-pantothenate, 1.0 g/L pyridoxine HCL, 1.0 g/L thiamine HCl and 25 mg/L minositol. To maintain a constant pH of around 5.6 the medium was buffered by adding 10 g/L succinic acid and 6 g/L NaOH. Glucose was autoclaved separately and the vitamin solutions was sterile filtered (pore size 0.2 µm Ministart®-Plus, Sartorius AG, Germany) and both added after autoclavation.

Shake flask cultivations

Pre-inocula of the engineered yeast diploid strains were incubated O/N in 3 mL of SC-his-ura medium with constant agitation of 150 rpm at 30°C. The shake-flask cultivations were performed in 500 mL flasks containing 50 mL of buffered mineral medium (pH=5.6), which were incubated at 30°C on an orbital shaker set to 150 rpm. The flasks were inoculated with an initial OD_{600} =0.01 and sampled at two time points of 48h and 60h. For comparison purposes only the final time point samples were considered.

Cell dry weight measurements

The biomass concentration was determined by cell dry-weight (DW) measurements using polyethersulfone (PES) filters with a pore size of 0.45 µm Montamil® (Membrane Solutions, LLC). The filters were pre-dried in a microwave oven at 150 W and weighed. 5 mL of cultivation broth was filtered and then washed with three volumes of distilled water. Finally, the filters with biomass were dried in the microwave oven at 150 W for 20 min and cooled down in a desiccator for a minimum of 2 hours. Finally, the filters with biomass were weighed and the cell DW was determined [23].

Extracellular metabolite measurements

Samples for the quantification of VG and its pathway catabolites were prepared as follows: 500 µL of fermentation broth and 500 µL of 96% EtOH were carefully mixed by vortexing and centrifuged at 12000xg for 2 min, the supernatant was transferred to a new tube and stored at -20 ^{o}C until further analysis. Extracellular vanillin β-D-glucoside (VG), vanillin (VAN), protocatechuic acid (PAC), protocatechuic aldehyde (PAL) and vanillic acid (VAC) were quantified using Agilent 1100 series equipment with a Synergi Polar-RP 150*2 mm 4u column (Phenomenex). A gradient of acetonitrile

(ACN) with 1% tetra-fluoroacetic acid (TFA) and water with 1% TFA at a constant flow rate of 0.5 mL/min was used as mobile phase. The elution profile was as follows: 5% ACN for 1 min, 5% ACN to 30% ACN for 8 min, 30% ACN to 100% ACN for 1 min, 100% ACN for 1 minute, 100% ACN to 5% ACN for 3 min. The column was kept at 40 ^{o}C and metabolite detection was performed using a UV diode-array detector set to 230 and 280 nm.

Results

Gene overexpression strategy

In this work, we decided to manipulate the dosage of the heterologous VG pathway genes by implementing the gene amplification (GA) system described in **Chapter 4**. The system relies on the use of specially designed yeast strains with predefined overexpression capacity, where gene amplification occurs in controlled manner by genomic integration into well characterized loci. As a starting point we have chosen the GA strain with an amplification capacity of seven copies, with three integration sites on chromosome X and four sites on XII [24].

To shorten the strain construction time, we have setup a fast and combinatorial method based on the use of yeast diploid cells (see **Figure 5.2**). First, the library of haploid strains carrying different parts of the amplified VG pathway was created (**Table 5.3**). Later, each of these haploids was independently crossed to the haploid of opposite mating type with the fully functional VG pathway resulting in the collection of prototrophic diploids carrying the VG pathway and seven additional copies of gene/s of the interest (GOI) (see **Materials and Methods**). In total more than 9x2 unique diploid strains were constructed for the VG pathway titration analysis (see **Table 5.4**).

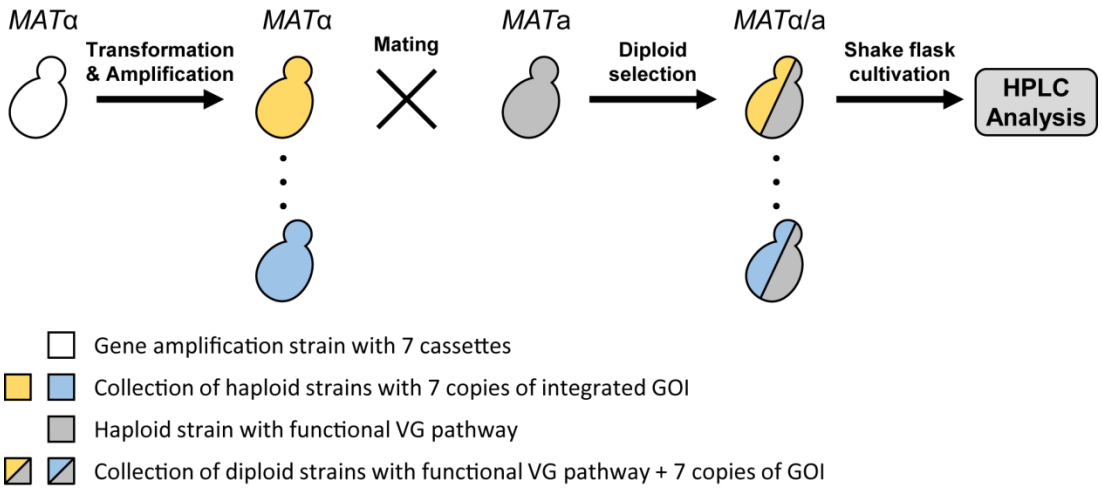


Figure 5.2. Schematic representation of the strategy used for overexpression of vanillin-β-glucoside pathway genes; GOI – gene/s of interest, VG – vanillin-β-glucoside.

Metabolic profile of the reference strain for VG production

All previous studies involving VG production in yeast were performed in haploid cells [5], [6]. However, diploid yeast are known to a have larger cell volume and the double amount of genetic material which might result in different VG production profile compared to the one obtained in haploid cells. Hence, to evaluate the effects of increased gene dosage of VG pathway genes, a proper diploid reference strain had to be developed. This was done by mating two specially constructed haploid strains (carrying double deletion of *adh6* and *exg1*), where only one of the mates had the VG pathway resulting in the heterozygous diploid strain D-VG-00 (see **Materials and Methods** and **Table 5.4**). The metabolic profile of the reference strain is shown in **Figure 5.3**, which is based on C-mol(metabolite)/C-mol(biomass) and the sum of all metabolites is normalized to 1 (arbitrary unit). The most pronounced metabolite appeared to be PAC (≈40 % of total metabolites), which is in line with previous observations by Brochado [5], [6]. The distribution of other VG pathway metabolites was somewhat correlated to the results obtained in **Chapter 6** and was as follows: PAL ≈2 %, VAC ≈24 %, VAL ≈5 % and VG ≈ 18 %. In addition to that, small quantities of isovanillin (IVAN) and isovanillic acid (IVAC) were detected (see **SI Table S 5.5**). To reduce the complexity of data interpretation, the side products IVAN and IVAC, representing less than 10% of the total metabolites, were neglected.

Overexpressing ACAR and its activating PPTase

Accumulation of the intermediates PAC and VAC in the reference strain D-VG-00 seems to be caused by limited activity of the ACAR. It is well known that ACARs as well as the related fatty acid synthetases, polyketide synthetases and non-ribosomal peptide synthetases require specific

activation by phosphopantetheinylation by PPTase enzymes [25], [26]. In this study, ACAR was activated by heterologous expression of bacterial PPTase *EntD,* which was found to perform the best compared to other PPTases tested [27]. Both, ACAR and EntD originate from different host organisms, and nothing is known about efficiency and interaction of those enzymes in *S. cerevisiae*. Thus, to investigate if the bottlenecks (PAC reduction to PAL and/or VAC reduction to VAN) are caused by low activity of ACAR or inefficient ACAR activation by PPTase, we constructed three different strains D-VG-20, D-VG-23 and D-VG-03 with overexpressed ACAR, ACAR+PPTase and PPTase, respectively. The distribution of the VG pathway metabolites produced by both D-VG-20 and D-VG-23 strains appeared to be very similar. In both strains, elevated ACAR activity resulted in 10 fold increased production of PAL and absent or significantly decreased accumulation (2.3 fold) of VAC (**Figure 5.3**). The VG production also improved by 2.1 fold in D-VG-20 and 2.7 fold in D-VG-23. As expected, experimental results of the PPTase overexpression, did not show significant changes in PAL and VAC levels.

Independent overexpression of OMT or UGT leads to a minor increase in VG production

A previous study had suggested OMT as a limiting step in VG production [6]. To confirm this, we have decided to check the effect of overexpressed OMT in the diploid cell based system. To our surprise, increased OMT dosage was not reflected in the metabolic profile of the D-VG-04 strain and minor improvement (1.6 fold) in VG production was detected (**Figure 5.3**). The distribution of the rest of the intermediates of the VG pathway were found to be similar to the ones found in the reference strain D-VG-00 (see also **Table S 5.5**).

The absence of VAN in the cultivation broth of the reference strain indicates that the UGT enzyme is capable to efficiently glycosylate VAN. However, we decided to test if overexpressed UGT would improve the flux towards VG production by creating metabolic “sink” by fast turnover of VAN to VG. The strain D-VG-50 indeed showed a 1.7 fold improvement in VG production (**Figure 5.3**). It can also be seen, that efficient VAN glycosylation made the VAC reduction by ACAR and PAL methylation by OMT more favorable, as the accumulated VAC and PAL were approximately 2 fold decreased. In addition to that, VAN reduction by the remaining activities of native yeast AHDs was outcompeted by overexpressed UGT, and therefore VAL was present in barely detectable quantities (**Table S 5.5**).

Coupled gene overexpression displays the highest impact on VG production

In addition to independent ACAR, OMT and UGT overexpression, we investigated the effect of combined overexpression of those genes. For this we constructed the strains D-VG-24, D-VG-25 and D-VG-54 harboring eight copies of ACAR+OMT, ACAR+UGT and UGT+OMT, respectively (**Table 5.4**). Notable changes in metabolic profiles of all three strains were observed (**Figure 5.3**). The strain D-

VG-54 displayed remarkably higher (4.1 fold compared to the reference) amount of produced VG, however no changes were observed for the remaining VG catabolites. On the other hand, the strains D-VG-24 and D-VG-25 had very similar PAL and VAC production profiles compared to the ones with overexpressed ACAR (D-VG-20 and D-VG-23). In both cases increased amount of PAL and decreased amount of VAC was observed see **Figure 5.3**. However, simultaneous overexpression of VG pathway steps 2 and 3 (**Figure 5.1**) in the strain D-VG-24 resulted in significant accumulation of VAN which in fact was a reason for increased amounts of VAL. The latter problem (accumulation of VAN) was absent in the strain D-VG-25 where amplified UGT helped to efficiently glycosylate VAN. Lastly, the production of VG was improved by four-fold in the D-VG-24 strain and as high as six-fold in the D-VG-25 strain.

Overexpressed 3DSD leads to increased flux towards VG pathway

In the VG pathway, PAC is formed by dehydration of the native precursor 3-DHS in a reaction catalyzed by 3DSD. Despite the fact that PAC is the most abundant intermediate, we decided to check if increased activity of the 3DSD would increase the carbon flux towards the heterologous pathway. Indeed, the strain D-VG-01 carrying eight copies of the *3DSD* gene produced 4.5 fold larger amount of PAC compared to the reference D-VG-00 (**Figure 5.3**). As expected, due to the basal activity of the remaining VG pathway enzymes no significant improvement was detected for the final product VG. However, overproduced PAC resulted in a slight increase of the products VAC and PAL of the following enzymatic steps carried by OMT and ACAR. Most importantly, the total amount of carbon redirected towards the *de novo* pathway increased by a factor of 2.7 (**Figure 5.3**), indicating highly improved potential for *de novo* production of VG.

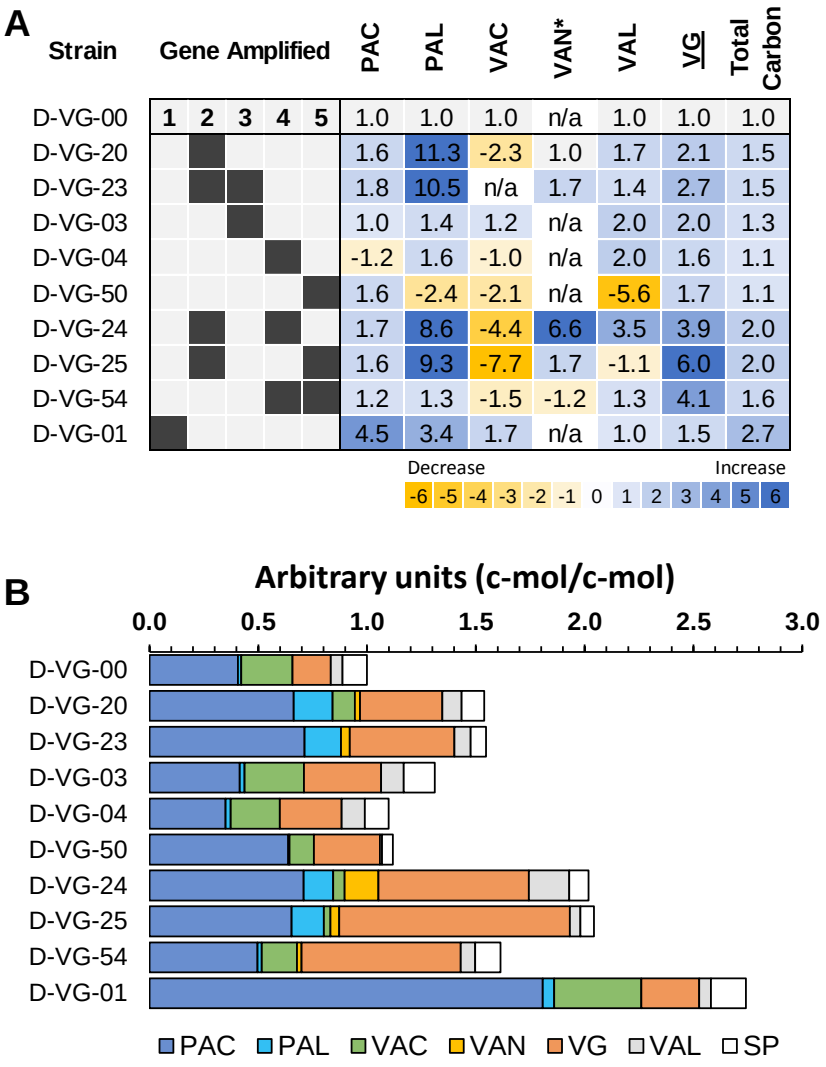


Figure 5.3 Results from the VG pathway gene overexpression. **A)** Represents the fold change of VG pathway metabolites in the strains with overexpressed gene/s (represented by black squares) with respect to the reference strain D-VG-00. *-for VAN metabolite, D-VG-20 was used as a reference. **B)** Metabolite distribution of the strains with overexpressed VG pathway genes, represented as c-mol(metabolite)/c-mol(biomass) and normalized to 1 according to the reference strain D-VG-00.

Discussion

Metabolic engineering of the *de novo* VG pathway was previously done by Brochado et al [6]. However, only two proteins of the pathway ACAR and OMT were overexpressed by genomic integration of an additional copy of either gene. In this study we have systematically overexpressed all VG pathway genes and several gene combinations by eight chromosomally integrated copies. We have demonstrated that by the increased dosage of different VG pathway genes it was possible to significantly improve the biosynthesis of VG and in some cases to reduce or eliminate the accumulation of unwanted intermediates (see **Figure 5.3**). It is worth to mention that, this study was based on an assumption that necessary cofactors (ATP, NADPH, SAM and UDP-Glucose) associated with the VG pathway enzymes are present/available in sufficient amounts and, therefore, previously described global metabolic engineering targets for increasing of the cofactor pools [5] were not pursued.

It appears that the *de novo* biosynthesis of VG in yeast is hampered by the wide substrate specificity of the heterologous pathway enzymes [1], creating two possible biosynthetic routes represented in steps **2** and **3** in **Figure 5.1**, which was confirmed by the metabolic profile of the reference strain D-VG-00. In fact, a significant percentage of accumulated PAC and VAC suggested that ACAR might be the primary bottleneck in the synthesis of VG. As mentioned earlier in this text, the ACAR in yeast is being expressed in the form of an apo-protein which requires posttranslational activation by PPTase [26]. In this work we have experimentally demonstrated that the ACAR protein was efficiently activated by phosphopantetheinylation, as the overexpressed PPTase did not affect the metabolite distribution of the VG pathway in the strain D-VG-03. However, when PPTase was overexpressed together with ACAR (D-VG-23), a slight improvement in VAC turnover to VAN and overall production of VG was observed, compared to when only ACAR was amplified (D-VG-20). This suggests that in case of further overexpression of ACAR an additional copy (or another, more efficient version [28]) of PPTase should be considered. Nonetheless, in both cases increased ACAR activity resulted in a buildup of the toxic intermediate PAL. Thus, overexpression of the O-methyltransferase encoded by OMT was also pursued to increase the metabolic flux from PAL to VAN. The additional dosage of OMT on top of the overexpressed ACAR (D-VG-24) did not affect the PAL dynamics, nevertheless, the flux through the pathway steps **2** and **3** was noticeably improved, represented by the accumulation of VAN and a fourfold increased amount of VG (**Figure 5.3**).

In previous optimization studies for VG production [5], [6] only one of the two possible biosynthetic routes (PAC→PAL→VAN→VG) was considered. However, based on the results presented above, we hypothesize that the VG biosynthesis might go primarily via VAC, rather than PAL, i.e. via the upper

route in steps **2** and **3** in **Figure 5.3**. Despite that the enzymatic properties of heterologous ACAR and OMT expressed in *S. cerevisiae* are not known, metabolic profiles of D-VG-20, D-VG-23, D-VG-24 and D-VG-04 suggest that ACAR might exhibit higher affinity towards VAC compared to PAC. The latter is also supported by the results of the *in vitro* characterization studies of ACAR from *Nocardia sp.* (the same isolate as used in this study), which demonstrated that out of a broad substrate range [26], vanillic acid (VAC) was found to be the most preferable substrate for ACAR [28]. On the other hand, it seems that the enzyme OMT preferably methylated PAC but not PAL as the overexpression of ACAR+OMT increased the flux through the PAC→VAC→VAN route, while PAL turnover remained intact **Figure 5.3**. In addition to the undefined substrate preferences of OMT, we speculate that the limited activity of overexpressed OMT might be a result of its cellular mislocalization or protein misfolding (**SI Supplementary note 1**). Furthermore, it was also found that the OMT enzyme also synthesized several unwanted isoforms of intermediates (**SI Figure S 5.2** and **Table S 5.5**), by methylation of the *para*-hydroxyl group of VAN and VAC. This side activity makes the OMT enzyme a troublesome point in the VG biosynthesis.

The intermediate VAN is known to be toxic for yeast [7] and it has been demonstrated that its presence in fermentation media in concentration as low as 3 mM causes growth inhibition [1]. In *S. cerevisiae*, VAN is mainly detoxified by its reduction to vanillyl alcohol (VAL) by the natural alcohol dehydrogenases (ADHs), in particular ADH6 [1]. The gene coding for ADH6 is deleted in the VG producing strain. However, accumulation of VAL indicates the presence of another minor enzymatic activity of the remaining ADHs (**SI Figure S 5.2**).This appears more evident in case of D-VG-24, were the increased flux towards VG is being constrained by insufficient activity of UGT leading to accumulation of VAN and consequently VAL. Thus, it is obvious that the native ADHs enzymes and heterologous UGT are competing for their substrate – VAN. Overexpression of the UGT (D-VG-50) eliminated VAL and also reduced the accumulation of VAC (see **Figure 5.3**). We speculate that increased activity of UGT generates some kind of "metabolic sink" around VAN which thermodynamically favors reactions catalyzed by enzymes upstream in the VG pathway. In the original study by Hansen *et al.* it was also reported that the accumulation of several upstream metabolites decreased after the glucosylation step was introduced into the vanillin producing strain [1]. In this study, the advantage of UGT overexpression appeared to be even more prominent when it was combined with the increased gene dosage of either OMT or ACAR (D-VG-54 and D-VG-25, respectively) which resulted in the best metabolic distribution observed, with significantly improved VG production as well (**Figure 5.3** and **SI Table S 5.5**).

We found that in all overexpression variants the accumulation of PAC was not reduced, which indicates that the activity of the following enzymes ACAR and OMT has to be enhanced even further. The production of PAC is predefined by the availability of the precursor 3-dehydroshikimic acid (3-DHS) and the efficiency of the enzyme 3DSD. Since, there are no cofactors associated with reaction of 3-DHS to PAC, it was hypothesized that the PAC dynamics are directly linked to the aromatic amino acid biosynthesis and, therefore, yeast growth. Indeed, two independent studies have demonstrated a very good correlation between biomass yield and PAC production during batch fermentation experiments [5] and **Chapter 6**. In this study, however, we were able to demonstrate that the amount of PAC can be significantly increased by overexpression of 3DSD as can be seen in the metabolic profile of D-VG-01 (**Figure 5.3** and **SI Table S 5.5**). This also noticeably in the improved metabolic flux towards the VG pathway which, in fact, increased the theoretical potential for VG production by 270%. One needs to add that in case of the balanced *de novo* pathway, further improvements could be achieved by increasing the metabolic flux through the *S. cerevisiae* aromatic amino acid biosynthesis pathway. This was successfully demonstrated by the deregulation of key nodes of the shikimate pathway (in particular ARO3 and ARO4 [29]), which resulted in more than 200-fold increased extracellular concentration of shikimate (the metabolite of 3-DHS) [30].

Concluding remarks and Future perspectives

The present study aimed at pinpointing the rate limiting steps in the vanillin-β-glucoside biosynthetic pathway by overexpressing each step of the heterologous pathway. The results of this work suggested that there is more than one bottleneck in the pathway and that a combined gene overexpression is needed. We demonstrated that the enzymes ACAR and HsOMT are the primary rate-limiting steps in the VG production. Nevertheless, an overexpressed ACAR or OMT had to be further complemented by an increased activity of UGT, which is the last step in the VG biosynthesis. In this study the best result was a 6-fold improvement in VG production, which was obtained by overexpressing the combination of ACAR+UGT. Moreover, based on observations, we have proposed that the VG biosynthesis predominantly follows the PAC→VAC→VAN→VG route rather than PAC→PAL→VAN→VG. In conclusion, this study showed that the existing *de novo* VG pathway in the yeast *S. cerevisiae* has a high potential for successful VG production, however, there is still room for improvement.

Future work on the development of the yeast cell factory for VG production should focus on finding the right ratios of ACAR, OMT and UGT, which can be easily done by implementing our existing GA

platform (**SI Figure S 5.4**). It is also of high importance to focus on improving the heterologous enzyme specificities and catalytic activities, which could be done both by using alternative genes or by targeted protein engineering [31] as recently demonstrated[32]. Finding or engineering an ACAR enzyme with exceptional specificity towards VAC and not PAC would eliminate the accumulation of the toxic intermediate PAL and presumably improve the flux towards the production of VG. Alternatively, it was shown that protein fusion can improve the flux through desired nods of a heterologous pathway [33]. Following our proposed VG biosynthetic route the overexpressed protein fusions of 3DSD-OMT and ACAR-UGT might solve the existing accumulation of the intermediates and increase the flux towards VG production. Lastly, the fact that heterologous enzymes are dependent on various cofactors must not be ignored and additional metabolic engineering strategies for increased cofactor supply might also be pursued [5], [6].

References

[1] E. E. H. Hansen, B. B. L. Møller, G. R. G. Kock, C. M. Bünner, C. Kristensen, O. R. Jensen, F. T. Okkels, C. E. Olsen, M. S. Motawia, and J. Hansen, "*De novo* biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (Saccharomyces cerevisiae).," *Appl. Environ. Microbiol.*, vol. 75, no. 9, pp. 2765–74, May 2009.

[2] N. J. Walton, M. J. Mayer, and A. Narbad, "Vanillin," *Phytochemistry*, vol. 63, no. 5, pp. 505–515, Jul. 2003.

[3] M. H. Priefert, "Biotechnological production of vanillin," pp. 296–314, 2001.

[4] B. Kaur and D. Chakraborty, "Biotechnological and molecular approaches for vanillin production: a review.," *Appl. Biochem. Biotechnol.*, vol. 169, no. 4, pp. 1353–72, Feb. 2013.

[5] A. R. Brochado, C. Matos, B. L. Møller, J. Hansen, U. H. Mortensen, and K. R. Patil, "Improved vanillin production in baker's yeast through in silico design.," *Microb. Cell Fact.*, vol. 9, no. 1, p. 84, Jan. 2010.

[6] A. R. Brochado and K. R. Patil, "Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering.," *Biotechnol. Bioeng.*, vol. 110, no. 2, pp. 656–9, Feb. 2013.

[7] A. Iwaki, S. Ohnuki, Y. Suga, S. Izawa, and Y. Ohya, "Vanillin Inhibits Translation and Induces Messenger Ribonucleoprotein (mRNP) Granule Formation in Saccharomyces cerevisiae: Application and Validation," *PLoS One*, vol. 8, no. 4, 2013.

[8] P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman, and C. J. Paddon, "Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 3, pp. E111–8, Jan. 2012.

[9] K.-K. Hong and J. Nielsen, "Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries.," *Cell. Mol. Life Sci.*, vol. 69, no. 16, pp. 2671–90, Aug. 2012.

[10] I.-K. Kim, A. Roldão, V. Siewers, and J. Nielsen, "A systems-level approach for metabolic engineering of yeast cell factories.," *FEMS Yeast Res.*, vol. 12, no. 2, pp. 228–48, Mar. 2012.

[11] F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, and B. a Halkier, "USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.," *Nucleic Acids Res.*, vol. 35, no. 7, p. e55, Jan. 2007.

[12] H. H. Nour-Eldin, B. G. Hansen, M. H. H. Nørholm, J. K. Jensen, and B. a Halkier, "Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments.," *Nucleic Acids Res.*, vol. 34, no. 18, p. e122, Jan. 2006.

[13] M. H. H. Nørholm, "A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering.," *BMC Biotechnol.*, vol. 10, p. 21, Jan. 2010.

[14] S. Partow, V. Siewers, S. Bjørn, J. Nielsen, and J. Maury, "Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae," *Yeast*, no. July, pp. 955–964, 2010.

[15] U. Güldener, S. Heck, T. Fielder, J. Beinhauer, and J. H. Hegemann, "A new efficient gene disruption cassette for repeated use in budding yeast.," *Nucleic Acids Res.*, vol. 24, no. 13, pp. 2519–24, Jul. 1996.

[16] S. Nakagawa, Y. Niimura, T. Gojobori, H. Tanaka, and K. Miura, “Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes.,” *Nucleic Acids Res.*, vol. 36, no. 3, pp. 861–71, Feb. 2008.

[17] D. R. Cavener and S. C. Ray, “Eukaryotic start and stop translation sites.,” *Nucleic Acids Res.*, vol. 19, no. 12, pp. 3185–92, Jun. 1991.

[18] D. Gietz, A. St Jean, R. a Woods, and R. H. Schiestl, “Improved method for high efficiency transformation of intact yeast cells.,” *Nucleic Acids Res.*, vol. 20, no. 6, p. 1425, Mar. 1992.

[19] D. a Treco and F. Winston, “Growth and manipulation of yeast.,” *Curr. Protoc. Mol. Biol.*, vol. Chapter 13, no. April, p. Unit 13.2, Apr. 2008.

[20] G. Bertani, “Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.,” *J. Bacteriol.*, vol. 62, no. 3, pp. 293–300, Sep. 1951.

[21] F. Sherman, G. R. Fink, and J. B. Hicks, *Laboratory course manual for methods in yeast genetics*. Cold Spring Harbor Laboratory, 1986.

[22] C. Verduyn, E. Postma, W. a Scheffers, and J. P. Van Dijken, “Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.,” *Yeast*, vol. 8, no. 7, pp. 501–17, Jul. 1992.

[23] L. Olsson and J. Nielsen, “On-line and in situ monitoring of biomass in submerged cultivations,” *Trends Biotechnol.*, vol. 15, no. December, pp. 517–522, 1997.

[24] M. D. Mikkelsen, L. D. Buron, B. Salomonsen, C. E. Olsen, B. G. Hansen, U. H. Mortensen, and B. A. Halkier, “Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform.,” *Metab. Eng.*, vol. 14, no. 2, pp. 104–11, Mar. 2012.

[25] A. He, T. Li, L. Daniels, I. Fotheringham, P. N. Rosazza, and J. P. N. Rosazza, “Nocardia sp . Carboxylic Acid Reductase : Cloning , Expression , and Characterization of a New Aldehyde Oxidoreductase Family Nocardia sp . Carboxylic Acid Reductase : Cloning , Expression , and Characterization of a New Aldehyde Oxidoreductase Family,” vol. 70, no. 3, pp. 1874–1881, 2004.

[26] P. Venkitasubramanian, L. Daniels, and J. P. N. Rosazza, “Reduction of carboxylic acids by Nocardia aldehyde oxidoreductase requires a phosphopantetheinylated enzyme.,” *J. Biol. Chem.*, vol. 282, no. 1, pp. 478–85, Jan. 2007.

[27] A. Brochado, *Investigating Genotype-phenotype Relationships in Saccharomyces Cerevisiae Metabolic Network Through Stoichiometric Modeling: Ph. D. Thesis*, no. April. Center for Microbial Biotechnology, Technical University of Denmark, 2012.

[28] P. Venkitasubramanian, L. Daniels, S. Das, A. S. Lamm, and J. P. N. Rosazza, “Aldehyde oxidoreductase as a biocatalyst: Reductions of vanillic acid.,” *Enzyme Microb. Technol.*, vol. 42, no. 2, pp. 130–7, Jan. 2008.

[29] S. Teshiba, R. Furter, P. Niederberger, G. Braus, G. Paravicini, and R. Hütter, “Cloning of the ARO3 gene of Saccharomyces cerevisiae and its regulation.,” *Mol. Gen. Genet.*, vol. 205, no. 2, pp. 353–7, Nov. 1986.

[30] M. a H. Luttik, Z. Vuralhan, E. Suir, G. H. Braus, J. T. Pronk, and J. M. Daran, “Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: quantification of metabolic impact.,” *Metab. Eng.*, vol. 10, no. 3–4, pp. 141–53, 2008.

[31] E. Nevoigt, “Progress in metabolic engineering of Saccharomyces cerevisiae.,” *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 3, pp. 379–412, Sep. 2008.

[32] J. HANSEN, E. H. HANSEN, H. P. SOMPALLI, J. M. SHERIDAN, J. R. HEAL, W. HAMILTON, and others, “COMPOSITIONS AND METHODS FOR THE BIOSYNTHESIS OF VANILLIN OR VANILLIN BETA-D-GLUCOSIDE.” 2013.

[33] L. Albertsen, Y. Chen, L. S. Bach, S. Rattleff, J. Maury, S. Brix, J. Nielsen, and U. H. Mortensen, “Diversion of flux toward sesquiterpene production in Saccharomyces cerevisiae by fusion of host and heterologous enzymes.,” *Appl. Environ. Microbiol.*, vol. 77, no. 3, pp. 1033–40, Feb. 2011.

[34] W. Huh, J. Falvo, L. Gerke, and A. Carroll, “Global analysis of protein localization in budding yeast,” *Nature*, vol. 425, no. 6959, pp. 686–91, Oct. 2003.

Supplementary Information for Chapter 5

Contents

Supplementary note 1

GFP tagging reveals plausible cellular localization of O-methyltransferase – one of the limiting steps in the VG biosynthesis

When expressing heterologous metabolic pathways in yeast, it is important that all enzymes are properly folded and localized in the same cellular compartment, most commonly the cytoplasm. The *de novo* Vanillin-β-glucoside (VG) pathway is assembled from enzymes originating from several evolutionary distant organisms, though, little is known about the actual expression of these proteins in *S. cerevisiae*. Misfolding, co-localization and aggregate formation of any protein from the VG pathway might lead to inefficient production and intermediate accumulation, which was observed by several studies [5], [6]. One way of checking the location of the protein in the cell is by means of fusing with a fluorescent reporter protein [34].

The aim of this experiment was to investigate whether all VG pathway enzymes are localized in the same cellular compartment, preferably the cytosol. In addition, fluorescence levels could indicate the relative expression levels of the VG pathway enzymes. To do that, we have C-terminally tagged all five heterologous proteins with either red (RFP) or cyan fluorescent proteins (CFP).

In this study, three strains with tagged VG pathway proteins were constructed see **Table S 5.1**. This was done by transforming the VG producing strain TS155 (one of the intermediate strains from the study described in **Chapter 6**) with DNA substrates shown in **Figure S 5.1A**. The tagged ORFs were integrated into defined location on the chromosome XI at position 2 [24]. Fluorescence microscopy of the generated strains revealed plausible localization of the OMT protein, while the remaining four were found in the cytoplasm (**Figure S 5.1B**). Moreover, the fluorescence signal of the OMT_RFP was relatively low compared to e.g. the one of the PPT_RFP fusion. The pattern of tagged OMT in the cell closely resembles the one of the proteins localized in the vacuole [34]. However, to determine the exact cellular location of the OMT, additional experiments need to be done. Firstly, to rule out if the localization is not an artifact of the tagging by RFP, the OMT should be also tagged N-terminally. Secondly, to confirm the exact cellular location of OMT, the proteins with known localization should be tagged simultaneously in the same strain. Alternatively, chemical staining of cell organelles can be done.

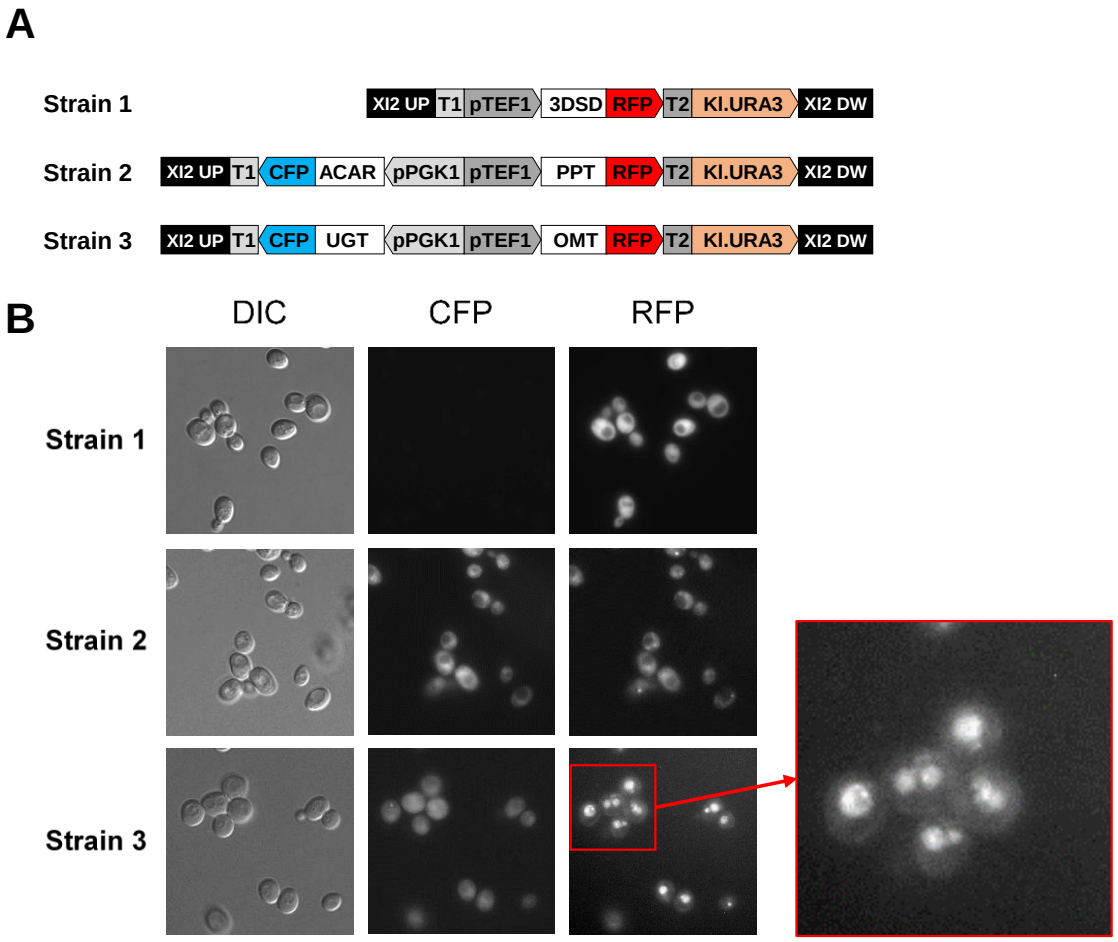


Figure S 5.1 Tagging proteins of vanillin-β-glucoside pathway proteins. **A)** Schematic representation of protein tagging. Molecular cloning was done using the methods as described in **Chapter 6. B)** Fluorescent microscopy results. To analyze the localization of the VG pathway proteins, strains 1) TS155-3R, 2) TS155-CAPR and 3) TS155-CUOR were incubated overnight in 20 mL of SC media with constant shaking 150 rpm and temperature set to 30°C.

Table S 5.1 List of the yeast strains used for the enzyme tagging experiment.

Strains	Relevant genotype	Reference
TS155	*MATα MAL2-8C SUC2 ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD), ΔEXG1::loxP ΔADH6::KanMX*	Chapter 6
TS155-3R	*MATα MAL2-8C SUC2 ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD), XI2(pTEF1-3DSD_RFP), ΔEXG1::loxP ΔADH6::KanMX*	This study
TS155-CAPR	*MATα MAL2-8C SUC2 ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD), XI2(pTEF1-PPT_RFP, pPGK1-ACAR_CFP), ΔEXG1::loxP ΔADH6::KanMX*	This study
TS155-CUOR	*MATα MAL2-8C SUC2 ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD), (pTEF1-HsOMT_RFP, pPGK1-UGT_CFP), ΔEXG1::loxP ΔADH6::KanMX*	This study

Table S 5.2 list of verification primer pairs used for the validation of gene targeting events. Appearance of the correct size band on the gel verifies targeted integration of your favorite fragment (XII-1, XII-2 and XII-5). In case of deletions, the band size confirms your favorite gene deletion. *In case of ADH6 the band size also includes the KanMX cassette.

Site	Name	Sequence	Fragment size, bp
XII-1	XII-1-up-out-sq	CTGGCAAGAGAACCACCAAT	729
XII-2	XII-2-up-out-sq	CGAAGAAGGCCTGCAATTC	739
XII-5	XII-5-up-out-sq	CCACCGAAGTTGATTTGCTT	688
—	C1_TADH1_F	CTTGAGTAACTCTTTCCTGTAGGTC	
For	*Deletion*	*Verification*	
BGL1	BGL1_VF	TCATCCTCCCTGTGTTTACA	1044
	BGL1_VR	AGTTGAAACAGAGGATAAGGTG	
*ADH6**	ADH6_VF	GTTTTGCTTTTTTCTCTGGG	2773
	ADH6_VR	GGAGTATCAACCACTAAAGCG	

Table S 5.3 list of verification primers pairs used for checking if the gene amplification has appeared. *Band sizes obtained when the reverse primer is C1_TADH1_F (**Table S 5.2**).

Site	Name	Sequence	Size, bp*
X-2	X-2-up-out-sq	TGCGACAGAAGAAAGGGAAG	1395
X-3	X-3-up-out-sq	TGACGAATCGTTAGGCACAG	1581
X-4	X-4-up-out-sq	CTCACAAAGGGACGAATCCT	1505
XII-1	XII-1-up-out-sq	CTGGCAAGAGAACCACCAAT	1374
XII-2	XII-2-up-out-sq	CGAAGAAGGCCTGCAATTC	1384
XII-3	XII-3-up-out-sq	TGGGCAGCCTTGAGTAAATC	1484
XII-4	XII-4-up-out-sq	GAACTGACGTCGAAGGCTCT	1350

Table S 5.4 the list of the additional primers used for sequencing purposes of the plasmids

Name	Sequence
S1_TEF_F	CGGTCTTCAATTTCTCAAGTTTC
S1_PGK_F	CAAGGGGGTGGTTTAGTTTAGT
S1_ACAR_F	CATTGTTCGTTCCAGACACTGAC
S2_ACAR_F	GAAATTGTCTCAAGGTGAGTTCG
S3_ACAR_F	TTTGGGTAGATTCTTGTGTTTGG
YF19	AAAAAATAAATAGGGACCTAGACTTCA
YF21	GACCTACAGGAAAGAGTTACTCAAGAAT

Table S 5.5 Concentration of the VG pathway metabolites, mg (g DW)$^{-1}$.

Strain	PAC	PAL	VAC	VAN	VG	VAL	IVAC	IVAN
D-VG-00	**43.49 ± 2.20**	**1.19 ± 0.18**	**23.86 ± 0.87**	**n/a**	**16.4 ± 2.9**	**5.2 ± 1.3**	**14.36 ± 0.17**	**0.11 ± 0.11**
D-VG-VG	101 ± 21	8.26 ± 0.63	26.4 ± 5.8	6.87 ± 1.35	72.4 ± 5.8	15.98 ± 1.10	10.91 ± 3.57	3.09 ± 0.10
D-VG-01	193.42 ± 0.04	5.15 ± 0.91	40.8 ± 0.2	n/a	29.1 ± 8.8	5.02 ± 0.63	15.89 ± 1.45	0.53 ± 0.53
D-VG-20	77.53	18.5	13.0	n/a	38.6	8.47	6.66	4.45
D-VG-03	44.3 ± 6.9	2.19 ± 0.03	27.8 ± 12.1	n/a	38.8 ± 6.7	9.75 ± 1.99	14.09 ± 4.40	0.43 ± 0.09
D-VG-04	37.4 ± 2.3	2.37 ± 0.09	22.9 ± 2.0	n/a	31.0 ± 2.0	10.11 ± 0.47	10.45 ± 0.07	0.62 ± 0.01
D-VG-50	68.21 ± 2.41	0.63 ± 0.12	11.4 ± 0.6	n/a	33.1 ± 2.7	0.88 ± 0.11	5.19 ± 5.19	n/a
D-VG-23	76.3 ± 24.2	16.0 ± 2.2	n/a	3.71 ± 2.10	52 ± 22	7.03 ± 2.02	4.58 ± 0.04	2.41 ± 0.70
D-VG-54	53.2 ± 4.2	1.91 ± 0.75	16.56 ± 1.68	1.82 ± 1.82	80 ± 25	6.25 ± 2.25	10.87 ± 1.06	0.84 ± 0.84
D-VG-00	**27.64 ± 4.22**	**1.16 ± 0.09**	**15.39 ± 2.71**	**n/a**	**13.85 ± 1.57**	**3.00 ± 0.34**	**5.50 ± 0.59**	**n/a**
D-VG-20	40.70	10.18	4.98	2.76	27.76	5.20	4.65	1.37
D-VG-24	48.05 ± 4.50	8.25 ± 1.85	3.45 ± 0.07	9.07 ± 0.40	47.81 ± 9.07	11.01 ± 0.97	2.78 ± 0.23	2.66 ± 0.20
D-VG-25	44.30 ± 5.54	8.98 ± 5.01	1.98 ± 0.21	2.39 ± 1.08	73.31 ± 6.05	2.84 ± 0.28	3.06 ± 0.24	0.91 ± 0.46

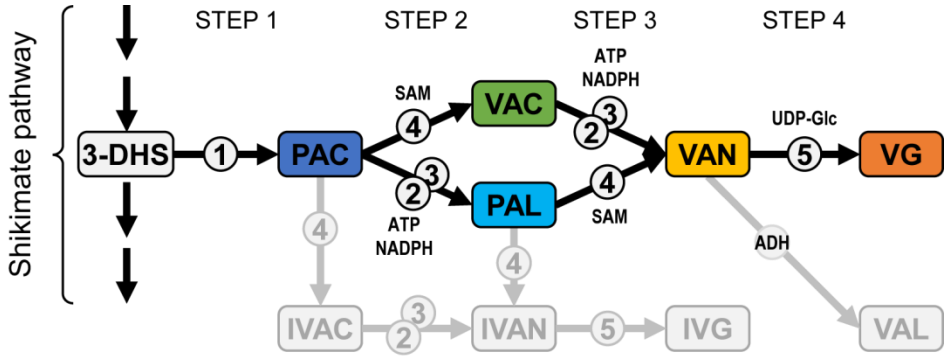


Figure S 5.2 Simplified heterologous pathway of vanillin-β-glucoside (VG) with side reactions shown. 3-DHS – 3-dehydroshikimic acid, PAC – protocatechuic acid, PAL – protocatechuic aldehyde, VAC – vanillic acid, VAN – vanillin, VG – vanillin-β-glucoside. Black thick arrows with numbered circles represent enzymatic reactions by heterologous enzymes: 1) 3DSD – 3-dehydroshikimate dehydratase (*Podospora pauciseta*), 2) ACAR – aromatic carboxylic acid reductase (*Neurospora sp.*), 3) EntD – phosphopantetheine transferase (*Escherichia coli*), 4) HsOMT – O-methyltransferase (*Homo sapiens*) and 5) UGT72E2 – UDP-glycosyltransferase (*Arabidopsis thaliana*). Grey arrows and boxes show unwanted side reactions; VAL – vanillyl alcohol, IVAC – isovanillic acid, IVAN – isovanillin and IVG isovanillin-β-glucoside. Cofactors: SAM – S-adenosyl methionine, ATP – adenosine triphosphate, NADPH – nicotinamide adenine dinucleotide phosphate.

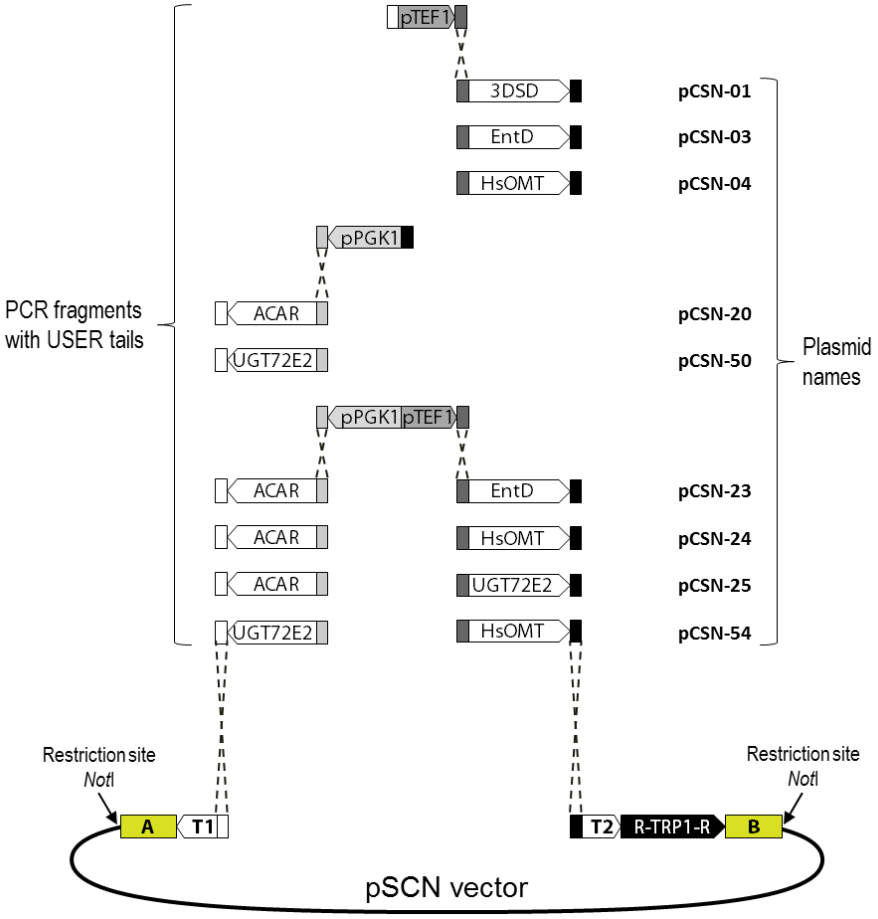


Figure S 5.3 Schematic representation of vector cloning for gene amplification purposes. The pCSN vector is linearized with AsiSI and Nb.BsmI enzymes. The yellow boxes A and B are targeting sequences homologous to the gene amplification cassette. T1 and T2 are terminator sequences Tadh1 and Tcyc1, respectively. The black arrow indicates the counter-selectable auxotrophic marker TRP1. White, grey and black boxes represent the USER cloning compatible tails.

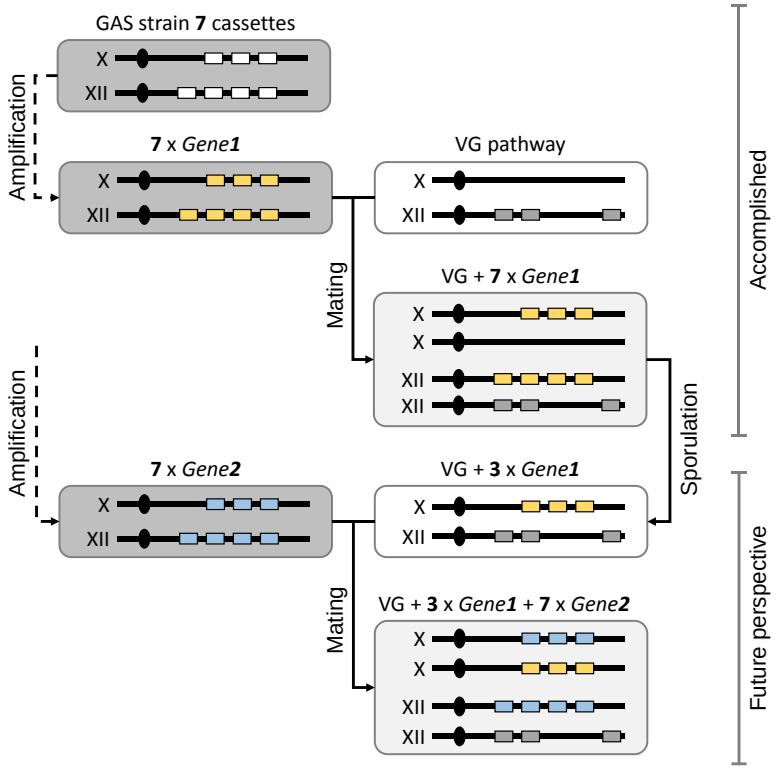


Figure S 5.4 Schematic representation of the VG pathway titration by implementing GAS strains. Each colorful box represents an integrated GOI as well as the location of the integration described by Mikkelsen (2012). By several rounds of genetic cross it is possible to generate a huge library with different combinations of amplified genes coding for VG pathway enzymes.

Chapter 6

Comparison of yeast strains for vanillin production

Title[6]

Benchmarking two commonly used *Saccharomyces cerevisiae* strains for heterologous vanillin-β-glucoside production

Abstract

The yeast *Saccharomyces cerevisiae* is a widely used eukaryotic model organism and a key cell factory for production of biofuels and wide range of chemicals. From the several available yeast strains, the most popular are the ones derived from laboratory strain S288c and the industrially relevant CEN.PK strain series. In recent years these strains have been subjected for comparison studies where genotypic and phenotypic differences were outlined. Based on high quality quantitative data obtained by multi-laboratory efforts it is hypothesized that strains S288c and CEN.PK might exhibit different biosynthetic capacity towards heterologous production. To test this, we have reconstructed previously published *de novo* vanillin-β-glucoside pathway, which uses an intermediate of aromatic amino acid biosynthesis (3-dehydroshikimate) as a precursor, in both S288c and CEN.PK strains in an identical manner. Characterization of the two producer strains revealed that the S288c background strain produced significantly higher amounts of vanillin-β-glucoside compared to CEN.PK in two standard growth conditions. This study demonstrated that yeast strain background may play a major role in the outcome of newly developed cell factories for production of a given product.

Keywords**:** Yeast; Cell factory, Strain choice; S288c, CEN.PK; Heterologous production; Vanillin-glucoside; Shikimate pathway.

[6] Manuscript in preparation

Introduction

With the recent achievements in the field of systems biology in combination with constantly advancing metabolic engineering and molecular biology tools, the yeast *Saccharomyces cerevisiae* is becoming a key cell factory for heterologous production of scientifically and industrially relevant products. The latter comprise a large variety of products ranging from low-value bulk chemicals and biofuels (e.g., ethanol) to food additives (e.g., flavors and colorants) and high value pharmaceuticals (e.g., recombinant proteins) [1], [2]. Today, a range of different *S. cerevisiae* strain backgrounds are available for the yeast community of which BY (S288c), W303 and CEN.PK are the most frequently used [3]. The variety of strains has been developed by different laboratories to suit a range of diverse research goals within different disciplines such as genetics, physiology and biochemistry. For example, CEN.PK strain, is a popular platform for physiological as well as metabolic engineering studies [3] whereas S288c, the first eukaryote to be sequenced, is mainly used for genetic studies, but has in recent years been increasingly used as an alternative platform for metabolic engineering experiments [4]. Specifically, recent surveys show that over the past ten years the two strains were used in more than 50% of the analyses with CEN.PK series being the most popular (approx. 37% - CEN.PK *vs.* 24% - S288c and its derivatives) [1], [2]. The importance of the two strains as cell factories is further substantiated by an extensive multi-laboratory efforts were made for systematic comparisons of S288c and CEN.PK [5]–[7]. Genetic differences of these strains were revealed first by microarray studies [8] and more recently genotype to phenotype relation was investigated after whole-genome sequencing of the CEN.PK113-7D strain [9], [10]. These sequencing studies revealed more than 13,000 single nucleotide polymorphisms (SNPs) in CEN. 113-7D compared to S288c, where 35% of those resulted in amino acid substitutions that affected 1406 proteins, moreover, 83 genes were absent in CEN.PK relative to S288C [10]. Considering only metabolic genes the highest SNPs enrichment was detected in the genes involved in the carboxylic acid, organic acid, and carbohydrate metabolism, as well as, nitrogen, amino acid, lipid and aromatic compound metabolism [9]. Several phenotypic differences were assigned to these mutations, e.g., differences in galactose uptake and ergosterol biosynthesis etc. in two strains.

The fact that the two main *S. cerevisiae* backgrounds for construction of cell factories are genetically/phenotypically quite different raises the possibilities that production yields also may be significantly different in CEN.PK and in S288c. To examine this possibility we have therefore constructed comparable cell factories for vanillin-β-glucoside (VG) production in S288c and CEN.PK backgrounds. Importantly, all five relevant genes for VG production in *S. cerevisiae*, (see Hansen et al. [12]) were integrated at phenotypically neutral and well defined locations in the yeast chromosome using a recently published integration platform [11]. Moreover, in order for the

engineered strains to be directly comparable, auxotrophies were eliminated by a sexual backcross to the corresponding wild-type strains. Remarkably, the physiological characterizations of both VG cell factories in two different cultivation modes revealed major differences in the VG production. Next, we examined the comprehensive –"omics" datasets for S288c and CEN.PK strains [6] to look for plausible reasons for the different VG production profiles in the two genetic backgrounds. Together our analyses serve as a step towards a scenario where the optimal genetic background for cell factory construction can be successfully selected based on a systems biology model for yeast cell factories.

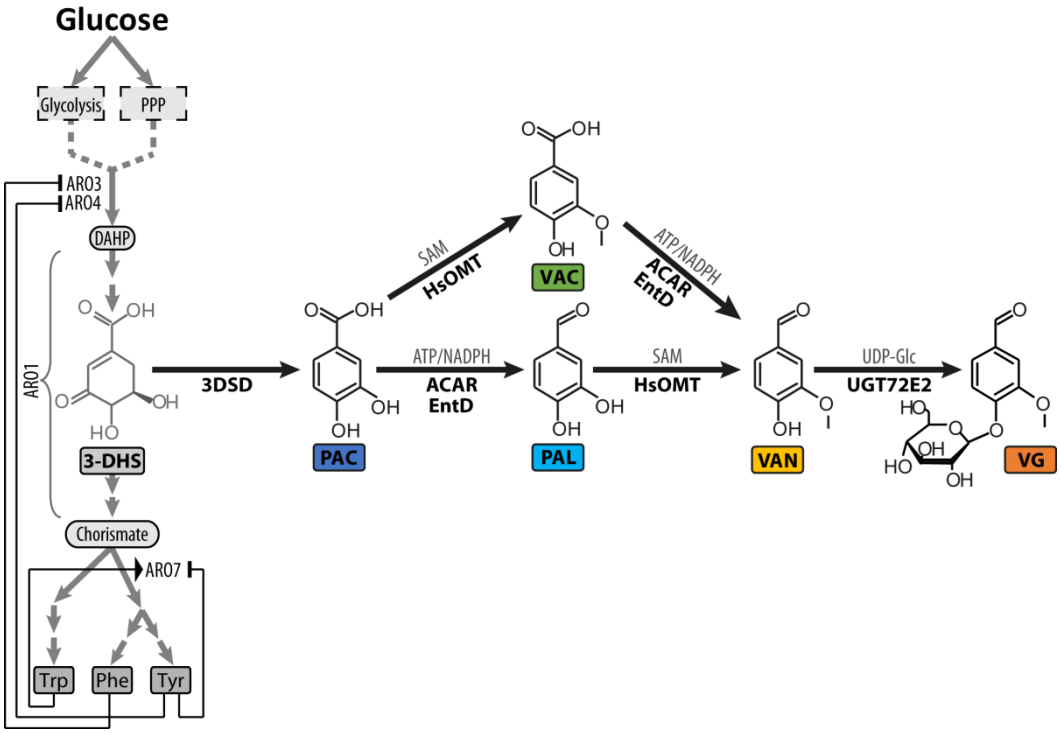


Figure 6.1. *De novo* vanillin-β-glucoside pathway in *S. cerevisiae* assembled by Hansen and co-workers [12]. Colored boxes represent metabolites of the pathway: PAC – protocatechuic acid, PAL – protocatechuic aldehyde, VAC – vanillic acid, VAN – vanillin, VG – vanillin-glucoside. Black thick arrows show enzymatic reactions by heterologous enzymes: 3DSD – 3-dehydroshikimate dehydratase (*Podospora pauciseta*), ACAR – aromatic carboxylic acid reductase (*Neurospora sp.*), EntD – phosphopantetheine transferase (*Escherichia coli*), HsOMT – O-methyltransferase (*Homo sapiens*) and UGT72E2 – UDP-glycosyltransferase (*Arabidopsis thaliana*). Grey arrows and boxes (left side of the picture) represent simplified shikimate biosynthetic pathway and native parts of yeast metabolism: PPP – pentose phosphate pathway, DAHP – 3-deoxy-D-arabinoheptulosonate 7-phosphate, 3-DHS – 3-dehydroshikimic acid, Trp – tryptophan, Phe – phenylalanine, Tyr – tyrosine. Key enzymes of the shikimate pathway: ARO3 and ARO4 – DAHP synthase isoenzymes and ARO1 – pentafunctional enzyme catalyzing DAHP conversion to chorismate.

Materials and Methods

DNA cloning procedures

The DNA fragments used for vector construction were amplified by PCR with PfuX7 polymerase developed by Nørholm et al. [13] using the primers listed in the **Table 5.2**. Molecular coning was done by uracil-specific excision reagent (USER™) as previously described in [14], [15].Final constructs were validated by sequencing (StarSEQ® GmbH, Germany). The genes constituting the *de novo* VG pathway (1-*3DSD*, 2- *ACAR*, 3-*EntD*, 4-*HsOMT*, and 5-*UGT72E2*) and a set of bidirectional promoters (pPGK1/pTEF1) were amplified by PCR from the appropriate vector templates (see **Table 5.1**). The pathway genes and promoters were assembled into vectors designed to integrate on chromosome XII [11]. A total of three plasmids were produced: pXII1-23 (pPGK1::ACAR; pTEF1::EntD), pXII2-54 (pPGK1::UGT72E2; pTEF1::HsOMT) and pXII5-01 (pTEF1::3DSD). For a full list of plasmids used or constructed in this work see **Table 5.1**; and for a schematic representation of plasmid cloning procedure refer to the **SI Figure S 6.1**.

Table 6.1. List of the plasmids used in this study.

Name	Genetic element	Reference
pXII1-23	*pPGK1::ACAR, pTEF1::EntD*	This study
pXII2-54	*pPGK1::UGT72E2, pTEF1::HsOMT*	This study
pXII5-01	*pTEF1::3DSD*	This study
pXII1	—	Mikkelsen et al. [11]
pXII2	—	Mikkelsen et al. [11]
pXII5	—	Mikkelsen et al. [11]
pSP-G2	*pPGK1;pTEF1*	Partow et al. [16]
pJH500	*3DSD*	Hansen et al. [12]
pJH674	*ACAR*	Hansen et al. [12]
pJH589	*EntD*	Hansen et al. [12]
pJH543	*HsOMT*	Hansen et al. [12]
pJH665	*UGT72E2*	Hansen et al. [12]
pUG6	*KanMX*	Güldener et al [17]
pSH47	*Cre recombinase*	Güldener et al [17]

Table 6.2. List of the primers used in this study. All sequences are presented in 5' to 3' direction, standard capital letter are gene specific sequences, **bold** underline letters represent USER specific tails, standard underline letters represent targeting sequences for appropriate gene deletions. *Italic* letters represents translational enhancer sequence [18], [19].

Name	Sequence
PGK_R-	ACCCGTTGAUGCCGCTTGTTTTATATTTGTTG
TEF_F+	CGTGCGAUGCCGCACACACCATAGCTTC
TEF_R+	ACGTATCGCUGTGAGTCGTATTACGGATCCTTG
DSD_F+	AGCGATACGU*AAAA*ATGCCTTCCAAACTCGCC
DSD_R+	CACGCGAUTTACAAAGCCGCTGACAGC
ACAR_F-	ATCAACGGGU*AAAA*ATGGCTGTTGATTCACCAGATG
ACAR_R-	CGTGCGAUCTTATAACAATTGTAACAATTCCAAATC
hOMT_F+	AGCGATACGU*AAAA*ATGGGTGACACTAAGGAGCAA
hOMT_R+	CACGCGAUCTTATGGACCAGCTTCAGAACC
PPT1_F+	AGCGATACGU*AAAA*ATGGTCGATATGAAAACTACGC
PPT1_R+	CACGCGAUTTAATCGTGTTGGCACAGC
UGT1_F-	ATCAACGGGU*AAAA*ATGCATATCACAAAACCACACG
UGT1_R-	CGTGCGAUACTAGGCACCACGTGACAAGTC
BGL1_del_F	ATTTTTTGTTTACTTTCTTTTTCTAGTTAATTACCAACTAAACT TCGTACGCTGCAGGTC
BGL1_del_R	CATTAGAAAATTCAGCTAAAATGAGCGGACTGAGGGCGAC TAGTGGATCTGATATCACCTA
ADH6_del_F	GAGGAAGAAATTCAACACAACAACAAGAAAAGCCAAAATC CTTCGTACGCTGCAGGTC
ADH6_del_R	GTTAAAAAGAAAGGAGCTACATTTATCAAGAGCTTGACAAC TAGTGGATCTGATATCACCTA

Strain construction

The genotype and source of the strains used in this study is given the **Table 5.3**. Two different background strains CEN.PK113-11C and X2180-1A (isogenic to S288c) were used as hosts for reconstruction of the *de novo* vanillin-β-glucoside (VG) pathway. All yeast constructs were generated by high efficiency transformation method (lithium acetate/polyethylene glycol/single carrier DNA) previously described by Gietz et al. [20].

To construct two yeast strains containing the VG pathway, vectors pXII1-23, pXII2-54 and pXII5-01 harboring the following genes (ACAR and EntD), (UGT72E2 and HsOMT) and (3DSD), respectively, were digested with *Not*I restriction enzyme (Fermentas-Thermo Fisher Scientific) and gel-purified using illustra GFX PCR DNA and Gel Band Purification Kit (GE Lifesciences). The individual gene targeting substrates were transformed iteratively into both yeast strain backgrounds in three consecutive transformations. To prevent undesired production of toxic intermediates especially protocatechuic aldehyde (PAL), the 3DSD gene was cloned at the latest step. After each round of

transformation the URA3 marker was eliminated by direct repeat recombination and counter-selection on 5-FOA [21] allowing the URA3 marker to be recycled.

Complete gene deletions of BGL1 and ADH6 were achieved using method described by Güldener et al [17]. PCR fragment carrying *loxP*-KanMX-*loxP* flanked by 40 nt long segment homologous to sequences of Up- and Down-stream of the appropriate open reading frame (ORF) to be deleted were amplified form plasmid pUG6. The KanMX marker was excised by expressing Cre recombinase from the vector pSH47 [17]. All gene targeting events were validated by diagnostic PCR using specially designed primer pairs (see **SI Table S 5.2**)

To eliminate any auxotrophies, two engineered strains S-VG-aux and C-VG-aux were sexually crossed to S288c and CEN.PK110-16D, respectively. The latter resulted in two final prototrophic vanillin-β-glucoside producing yeast strains S-VG (s288c based) and C-VG (CEN.PK based). Genetic cross and selection procedures were performed as described in [22].

Schematic flowchart representing the strain construction is depicted in **Figure S 6.2**.

Table 6.3. List of the yeast strains used in this study.

Name	Genotype	Reference
X2180-1A (ura3)	*MATα SUC2 mal mel gal2 CUP1 ura3-52*	Public domain [23]
S288c	*MATα SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6*	ATCC 204508
CEN.PK113-11C	*MATα MAL2-8C SUC2 ura3-52 his3Δ*	Peter Kötter[7]
CEN.PK110-16D	*MATα MAL2-8C SUC2 trp1-289*	Peter Kötter[1]
S-VG-aux	*MATα MAL2-8C SUC2 ura3-52 his3Δ XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	This study
C-VG-aux	*MATα SUC2 gal2 mal mel ura3-52 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	This study
S-VG	*MATα MAL2-8C SUC2 XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	This study
C-VG	*MATα SUC2 gal2 mal mel XII2(pTEF1-HsOMT, pPGK1-UGT) XII1(pTEF1-PPT, pPGK1-ACAR) XII5(pTEF1-3DSD) Δbgl1::loxP Δadh6::KanMX*	This study

Media

For cloning purposes lysogeny broth (LB) [24] supplemented with 100 mg/L of ampicillin was used for growing of *Escherichia coli* DH5α.

All media used for genetic manipulations of yeast were prepared as previously described by Sherman et al. [25], with minor modifications of synthetic medium where leucine concentration

[7] Institut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

was doubled to 60mg/L. All yeast transformants with gene integrations were selected on synthetic complete media without uracil. For subsequent rounds of transformations, the URA3 marker was recycled through direct repeat recombination and selected on synthetic complete media containing 30 mg/L uracil and 740mg/L 5-fluoroorotic acid (5-FOA) (Sigma-Aldrich).

Yeast transformants with necessary gene deletions were selected on Yeast Extract Peptone Dextrose (YPD) plates supplemented with 200 mg/L of G418 (Sigma) [17]. The medium composition was as follows: 10 g/L of yeast extract, 20 g/L of peptone and 20 g/L of glucose and 20 g/L of agar.

A defined minimal medium previously described by Verduyn [26] with 7.5 g/L glucose as a carbon source was used for *S. cerevisiae* cultivations in batch and chemostat cultures. Prior inoculation of fermenters, the medium was supplemented with 30 g/L of glucose [6]. The medium is composed of: 7.5 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.75 g/L Mg_2SO_4, 1.5 mL/L trace metal solution, 1.5 mL/L vitamin solution, 0.05 mL/L antifoam 204 (Sigma-Aldrich). Trace metal solution contained 3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 4.5 g/L $CaCl_2 \cdot 6H_2O$, 0.84 g/L $MnCl_2 \cdot 2H2O$, 0.3 g/L $CoCl_2 \cdot 6H_2O$, 0.3 g/L $CuSO_4 \cdot 5H_2O$, 0.4 g/L $NaMoO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 0.1 g/L KI and 15 g/L $Na_2EDTA \cdot 2H_2O$. Vitamin solution included 50 mg/L d-biotin, 200 mg/L para-amino benzoic acid, 1.0 g/L nicotinic acid, 1.0 g/L Ca-pantothenate, 1.0 g/L pyridoxine HCL, 1.0 g/L thiamine HCl and 25 mg/L minositol. Glucose was autoclaved separately and the vitamin solution was sterile filtered (pore size 0.2 μm Ministart®-Plus, Sartorius AG, Germany) and added after autoclavation.

Batch and chemostat cultivations

For each biological replicate, separate colonies of engineered yeast strains from YPD plate were inoculated to 0.5 L shake flasks with 50 mL of the previously described minimal medium (pH 6.5). Pre-cultures were incubated in an orbital shaker set to 150 rpm at 30°C until mid-exponential phase $OD_{600} \approx 5$ and directly used for inoculation. In this work, batch cultivations were performed in duplicates and continuous cultures – in triplicates.

Batch cultivations were performed under aerobic conditions in 1L fermenters equipped with continuous data acquisition (Sartorius, B. Braun Biotech International, GmbH, Melsungen, Germany) with a working volume of 1L. Fermenters were inoculated with initial $O.D._{600}$ =0.05. To ensure adequate mixing and aeration a stirrer speed was set to 600 r.p.m and airflow rate to 1.0 v.v.m. (60 L/h). The temperature was maintained at 30°C during the cultivation and pH=5.0 level was controlled by automatic addition of 2M NaOH or 2M H_2SO_4. The batch cultures were sampled at regular intervals through both glucose and ethanol growth phases. Glucose-limited chemostat cultures were grown in the same conditions as previously described for batch cultivations. Chemostats were initiated as batch cultivations with starting glucose concentration of 15 g/L and

OD_{600} =0.05 and switched to continuous mode in early exponential phase. Minimal medium was fed at a constant dilution rate of 0.1 h^{-1} [6].The working volume of 1 L was kept constant by a level based outlet. Samples were taken overr at least five residence times (50 hours) of constant biomass and carbon dioxide concentration readings.

In both cultivation experiments exhaust gas composition was constantly monitored by an off gas analyzer (1311 Fast response triple gas, Innova) combined with Mass Spectrometer (Prima Pro Process MS, Thermo Fisher Scientific, Winsford UK).

The biomass concentration was monitored by measuring both optical density at 600nm wavelength (OD_{600}) and cell dry weight (DW) in the cultivation broth. OD_{600} was estimated using a UV mini 1240 spectrophotometer (Shimadzu, Kyoto, Japan), biomass samples were diluted with distilled water to achieve OD_{600} reading within 0.1 to 0.4 range. DW measurements were performed using polyethersulfone (PES) filters with a pore size of 0.45 µm Montamil® (Membrane Solutions, LLC). The filters were pre-dried in a microwave oven at 150 W and weighed. A known volume of cultivation broth (5 mL) was filtered and then washed with approx. 15 mL of distilled water. Finally, the filters with biomass were dried in the microwave oven at 150 W and cell DW was determined [27].

Extracellular metabolite measurements

External metabolites were determined by high performance liquid chromatography (HPLC) analysis. Two distinct HPLC methods were applied for analysis of different groups of extracellular compounds.

The submerged cultivation samples for yeast primary metabolites were centrifuged at 12000xg for 2 min, supernatant was transferred to a new tube and stored at -20 °C until further analysis. Glucose, ethanol, glycerol, pyruvate, succinate and acetate were determined by high performance liquid chromatography (HPLC) analysis using an Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories, Hercules, CA). The column temperature was kept at 60 °C and the elution was performed using 5 mM H_2SO_4 with constant flow rate of 0.6 mL min^{-1}. Metabolite detection was performed by a RI-101differential refractometer detector (Shodex) and an UVD340U absorbance detector (Dionex) set at 210 nm.

Samples for quantification of VG and its pathway metabolites were prepared as follows: 1mL of fermentation broth and 1mL of 96% EtOH was carefully mixed by vortex and centrifuged at 12000xg for 2 min, supernatant was transferred to a new tube and stored at -20 °C until further analysis. Extracellular vanillin-β-glucoside (VG), vanillin (VAN), protocatechuic acid (PAC), protocatechuic

aldehyde (PAL) and vanillic acid (VAC) were quantified using Agilent 1100 series equipment with a Synergi Polar-RP 150*2 mm 4u column (Phenomenex). A gradient of acetonitrile (ACN) with 1% tetra-fluoroacetic acid (TFA) and water with 1% TFA at a constant flow rate of 0.5 mL/min was used as mobile phase. The elution profile was as follows: 5% ACN for 1 min, 5% ACN to 30% ACN for 8 min, 30% ACN to 100% ACN for 1 min, 100% ACN for 1 minute, 100% ACN to 5% ACN for 3 min. The column was kept at 40 ^{o}C and metabolite detection was performed using a UV diode-array detector set to 230 and 280 nm.

Intracellular metabolite measurements

Quenching and extraction: Samples for intracellular metabolites were taken in the end of continuous cultivation (during steady state conditions). The samples were quenched and extracted as previously described by Villas-Boas et al; 5 mL of culture broth was sprayed into pre-cooled (-40°C) falcon tube containing 20 mL of 60% methanol, spun down for 2 min at 5000xg in precooled centrifuge (-10°C) and extracted using boiling ethanol method [28] with addition of internal standards described below.

Stock solutions: Single stock solutions with concentration of 1 mg/ml of ATP, ADP, AMP, NADPH, NADP, NADH, NAD, UDP, UDP-glucose and shikimic acid were prepared in water and stored under –20°C until use. Aliquots of the stock solutions were used to prepare the daily working solutions by further dilution in 10 mM TBA and 10 mM acetic acid solution. Stock solutions (1 and/or 0.2 mg/ml) of labeled internal standards of [U-13C]ATP, [U-15N]AMP and [U-15N]UTP were prepared in water and kept under –20 °C until use. Mixture containing 30 µg/ml of [U-13C]ATP and 10 µg/ml of both [U-15N]AMP and [U-15N]UTP was used for spiking the calibration standards and the extracts.

Ion-pair UHPLC-MS/MS analysis of C-di-GMP. The analysis was carried out on an Agilent 1290 binary UHPLC system coupled with an Agilent 6460 triple quadrupole mass spectrometer (Torrance, CA, USA). The MS was operated in negative ion and multiple reactions monitoring mode. Separation of 0.5 µL samples was performed by ion-pair chromatography, as described in details in [29] using 10 mM tributylamine as ion pair reagent. The gradient used was 0-12 min 0-50% B, 12-12.5 min 50-100% B, 12.5-14 min 100% B, 14-14.5 min 100-0% B, 14.5-19.5 min 0% B. Calibration curves were constructed by preparing calibration solutions ranging from 1 to 100 µg/ml for ATP, NAD and UTP and 0.3 to 25 µg/ml for NADPH, NADP, NADH, NAD, UDP, UDP-glucose and shikimic acid. Both the extracts after the quenching and the calibration solutions were spiked with 60 µl of isotope labeled standard mixture. The calibration curves, for the compounds for which isotope-labeled analog was added, were constructed by plotting the peak area ratios of the unlabeled nucleotide standard to the corresponding labeled one versus the concentrations of the nucleotide standard added. For the

compounds where no IS was added, the calibration curves were constructed by plotting the peak area of the compound versus the concentration. The quantification was carried out using Mass Hunter Quantitative analysis software (version B.06.00).

Results

De novo pathway reconstruction in CEN.PK and S288c backgrounds

To compare CEN.PK and S288c for their ability to produce VG in a fair manner, it is important that all genes are inserted identically in both strain backgrounds. In the original VG producing strain the individual genes of the pathway were inserted by repeated integrations into the *TPI1* promoter region in an S288c based strain [12] and the gene order is consequently ill defined and therefore not easy to reconstruct in CEN.PK. We therefore introduced the VG pathway in the two strain backgrounds, S288c (isogenic isolate X2180-1A [23]) and CEN.PK113-11C [4], in a different manner. Specifically, the five genes used by Hansen et al.[12] for VG production were integrated into three sites located on chromosome XII, which are part of a defined gene expression platform we have previously established [11]. Importantly, prior to integration, we compared the up- and downstream sequences at these integration sites in CEN.PK and S288c and found that they differ by only a single SNP, a C (CEN.PK) to G (S288c) substitution in the upstream targeting sequence of XII-5 in genome (**SI Table S 6.3**). The five genes are therefore inserted into a genetic context, which is essentially identical. Finally, for both genetic backgrounds, auxotrophic markers were eliminated by sexual back crossing to wild-type variants of the two strain backgrounds to produce two prototrophic VG producing strains, C-VG (CEN.PK based) and S-VG (s288c based), which we used for further analysis.

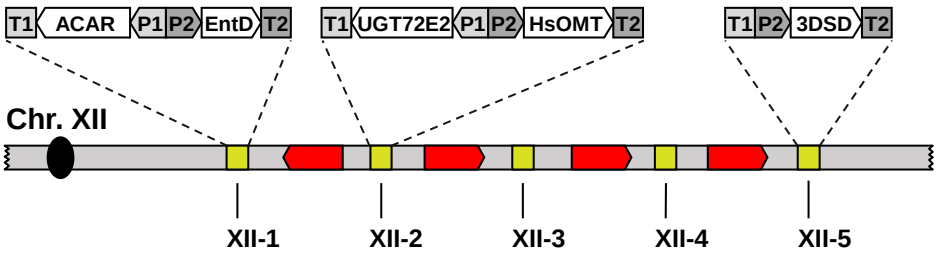


Figure 6.2. Schematic representation of VG pathway reconstruction in two different *S. cerevisiae* background strains. T1, T2 – terminators of *ADH1* and *CYC1*, respectively; P1, P2 – promoters of *PGK1* and *TEF1*, respectively. Red arrows represent essential genetic elements of *S. cerevisiae* and yellow boxes represent integration sites on chromosome XII characterized by Mikkelsen et al [11].

Physiological characterization of vanillin glucoside producing strains

Before evaluating the VG production ability of the two strain backgrounds, we first assessed whether VG production affected the overall physiology of the C-VG and S-VG strains. Hence, they were grown in batch and as continuous cultures in well-controlled bioreactors under standard cultivation conditions. Like for wild-type strains, the overall growth profiles exhibited by the S-VG and C-VG strains in batch reactors were composed by two growth phases. One initial growth phase where all glucose was fermented; followed, after the diauxic shift, by a second growth phase where all accumulated ethanol was respiro-fermented, see **Figure 1.3**. The main physiological parameters derived from these two experiments are summarized in **Table 6.4**.

Using identical growth conditions, a specific biomass yields on glucose (Y_{sx}) were found to be slightly higher in WT CEN.PK (~0.16 c-mol/c-mol) compared to the WT s288c-(~0.14 c-mol/c-mol) in previous studies [3], [6]. The same trend was found to be similar to what we find for C-VG and S-VG, respectively, in this experiment. Further analysis of the cultivation broth showed that the production profiles of five primary metabolites in the central carbon metabolism of S-VG and C-VG did not show any abnormalities (see **SI Figure S 6.4**). It has previously been observed that CEN.PK grows faster than S288c at these conditions, $\mu_{max} \approx 0.4\ h^{-1}$ vs. $\mu_{max} \approx 0.3\ h^{-1}$, respectively [3], [6], [9]. Importantly, we find that both VG producers displayed decreased and identical growth rates on glucose ($\mu_{max} \approx 0.2\ h^{-1}$) compared to the corresponding wild-type strains. Finally, we note that S-VG displayed a growth deficiency during the ethanol respiration phase, which for that reason is much longer for S-VG (23 h) than for C-VG (12 h). This deficiency has also been reported for wild-type S288c [9].

In chemostats, steady-state conditions with constant production of biomass as well as stable readings of carbon dioxide and oxygen by the off gas analyzer where obtained for both strain backgrounds. For the C-VG strain this was achieved in less than 5 residence times (50 hours) after feeding was initiated, whereas for the S-VG strain it took more than 8 residence times (80 hours), see **SI Figure S 6.3**. Importantly, at this dilution rate, both strains propagated exclusively by respiratory metabolism [30] as no production of ethanol, glycerol and organic acids was observed (see **Table 6.4)** The two strains produced similar amounts of biomass, and like in the batch fermentation experiment, the glucose uptake rate was slightly higher (7 %) in S-VG compared to C-VG.

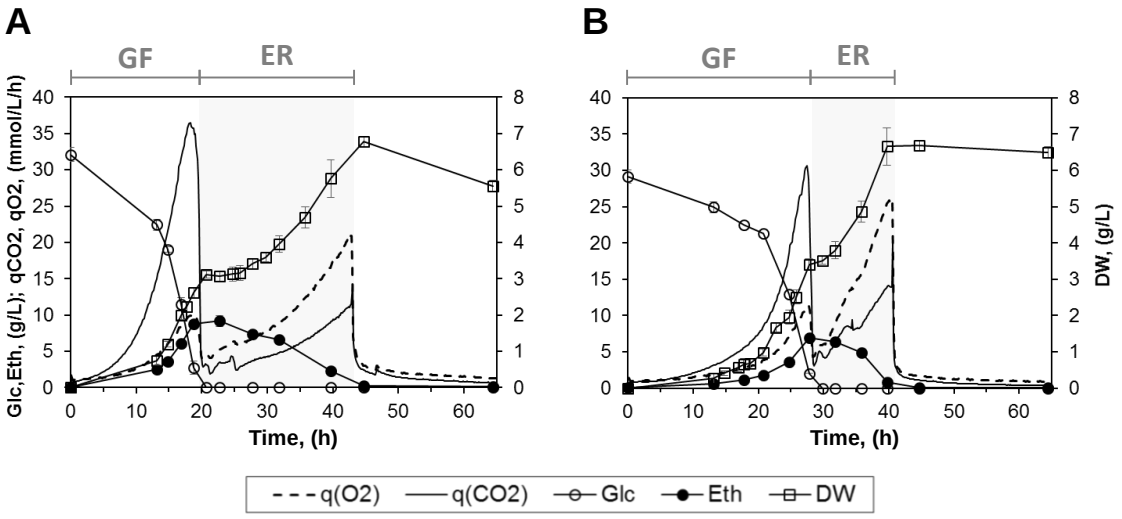


Figure 6.3. Growth profiles of two strains in the batch cultures: A) S-VG and b) C-VG. Grey areas/zones represent ethanol respiration (ER) phase; GF stands for glucose fermentation. qCO_2 – carbon dioxide production rate, qO_2 – oxygen consumption rate; Glc – glucose, Eth – ethanol and DW – dry weight.

Table 6.4. Physiologic parameters of the two engineered strains. Ysx – biomass yield on glucose, r – specific metabolite production or consumption rates (C-mmol/g(DW)·h); Glc –glucose, Eth – ethanol and Gly – glycerol. NA – not applicable and ND – not detected. The yield coefficient Y_{sx} for biomass is calculated based on a molecular weight for biomass of 26.4 g $Cmol^{-1}$ [31].

	Batch		**Chemostat**	
Strain	**S-VG**	**C-VG**	**S-VG**	**C-VG**
Glucose **μ**, h^{-1}	0.21 ± 0.01	0.20 ± 0.00	(0.1)*	(0.1)*
Ethanol **μ**, h^{-1}	0.05 ± 0.01	0.08 ± 0.00	NA	NA
$\mathbf{Y_{sx}}$ (C-mol/C-mol)	0.133 ± 0.014	0.151 ± 0.001	0.56 ± 0.02	0.60 ± 0.02
$\mathbf{-r_{Glc}}$	60.7 ± 8.0	49.9 ± 0.6	6.81 ± 0.29	6.37 ± 0.20
$\mathbf{r_{Eth}}$	25.9 ± 3.1	20.9 ± 3.0	ND	ND
$\mathbf{r_{Gly}}$	4.05 ± 0.05	1.40 ± 0.13	ND	ND

* - the dilution rate used in this study.

Metabolic profiles of S-VG and C-VG reveal significant differences in VG production in batch cultures

To compare VG production in the S-VG and C-VG strains, samples were taken at regular time points throughout the entire batch cultivation experiment. Remarkably, HPLC analysis of extracellular metabolite levels in the cultivation broth revealed obvious differences between the two VG producers. At the end of the experiment (t=45 hours), strain S-VG produced almost twice the amount of VG (39 mg/g_{Glc}) as compared to C-VG (20 mg/g_{Glc}). As has been previously observed

before, various levels of VG intermediates PAC, PAL, VAC and VAN were also produced at this time point due to inherently unbalanced heterologous VG pathway [12], [32]. Notably, despite that the two strain produce VG at very different levels the amounts of these intermediates were almost identical in the S-VG and C-VG strains (**Figure 6.4A**).Specifically, the metabolite PAC was the most abundant (close to 40 mg/g_{Glc} in both strains), whereas the amounts of other VG pathway metabolites PAL, VAC, and VAN, were significantly lower (ranging from 1.6 to 9 mg/g_{Glc}).

To account for the metabolic flux towards VG pathway in each strain, we have calculated total carbon based on the sum of all produced VG metabolites expressed in C-mmol/g_{DW} .It appeared that total carbon produced by VG pathway was 35 % higher in S-VG than in C-VG (**Figure 6.4B**). The final conversion to VG was as high as 43% in S-VG compared to only 27% in C-VG of total metabolites produced by integrated pathway.

Heterologous production of VG and its intermediates depends on yeast growth state in the batch cultivation

By monitoring the evolution of the VG pathway metabolites throughout both growth stages – glucose fermentation (GF) and ethanol respiration (ER) phase, differences in production profiles of two strains were observed (**Figure 6.5**c and SI **Figure S 6.5**). During the GF phase C-VG exhibited fairly better performance towards VG production which can be seen in **Table 6.5**. Here, the VG specific yield on glucose during exponential growth phase was approx. 37 % higher for C-VG compared to S-VG and was in correlation with the specific production rates (0.50 *vs.* 0.45 C-mmol/(g_{DW}·h), respectively). On the contrary, during ER phase the situation had reversed and the strain S-VG produced twofold more of VG compared to C-VG strain (**Figure 6.5**) in correlation with increased time for ethanol respiration by S-VG and slightly different volumetric productivities of (41 *vs.* 34 mg/(L·h), respectively. Nevertheless, other VG metabolites were produced in relatively similar manner by both strains. Most importantly, however, the total metabolic flux towards the VG pathway was lower in C-VG compared to S-VG in either GF or ER growth phases (**Figure 6.5**).

Unlike the other VG pathway intermediates, production of the PAC was not associated with any cofactors and was dependent on the availability of the precursor 3-DHS (**Figure 6.1**). In batch cultivation we have observed close correlation between PAC and biomass production in both strains during both GF and ER growth phases. However, it is clear from the **Figure 6.5A** and **B** that strain C-VG produced significantly less PAC per biomass compared to S-VG.

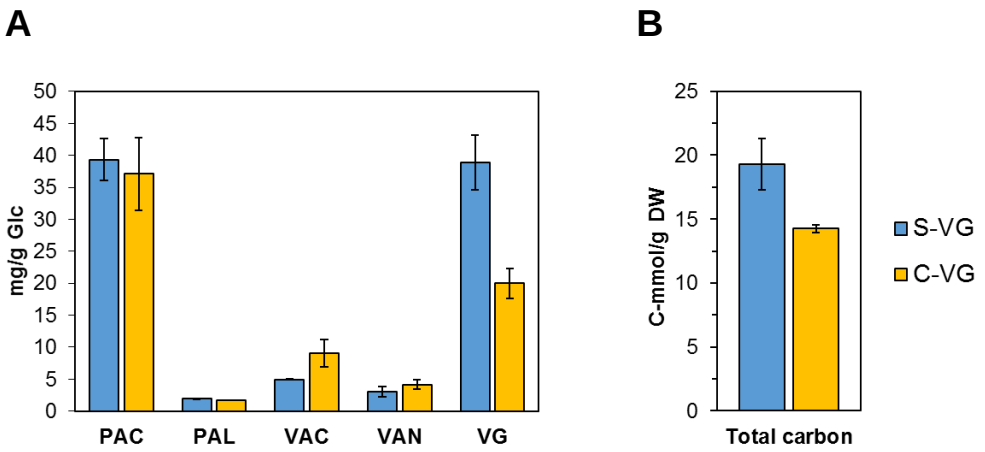


Figure 6.4. Final yields of VG and its intermediates produced by S-VG and C-VG during batch cultivation; A) based on final samples taken 45 hours after inoculation, B) sum of all detected VG metabolites normalized to dry-weight.

Table 6.5. Yields - $Y_{S\ Met}$ (mg/g(Glc)) and production rates of VG and its intermediates - r_{Met} (C-mmol/g(DW)·h).

	Batch [A]		**Chemostat**	
Strain	S-VG	C-VG	S-VG	C-VG
$\mathbf{Y}_{S\ PAC}$	19.0 ± 1.8	15.2 ± 0.2	3.57 ± 0.62	0.26 ± 0.04
$\mathbf{Y}_{S\ PAL}$	1.9 ± 0.2	2.0 ± 0.2	ND	0.09 ± 0.01
$\mathbf{Y}_{S\ VAC}$	1.5 ± 0.2	4.3 ± 1.5	0.88 ± 0.18	3.65 ± 0.09
$\mathbf{Y}_{S\ VAN}$	0.1 ± 0.1	0.2 ± 0.1	0.18 ± 0.14	ND
$\mathbf{Y}_{S\ VG}$	**5.4** ± 0.8	**7.4** ± 0.3	**13.7** ± 1.4	**1.06** ± 0.18
$\mathbf{r}_{PAC}$	1.59 ± 0.35	1.03 ± 0.03	0.64 ± 0.11	0.044 ± 0.008
$\mathbf{r}_{PAL}$	0.18 ± 0.04	0.15 ± 0.01	ND	0.016 ± 0.002
$\mathbf{r}_{VAC}$	0.13 ± 0.03	0.31 ± 0.10	0.16 ± 0.03	0.62 ± 0.03
$\mathbf{r}_{VAN}$	0.01 ± 0.01	0.01 ± 0.01	0.17 ± 0.12	ND
$\mathbf{r}_{VG}$	**0.45** ± 0.12	**0.50** ± 0.02	**2.47** ± 0.25	**0.18** ± 0.03

Note: A – for the batch experiments calculations were made only for exponential growth phase on glucose.

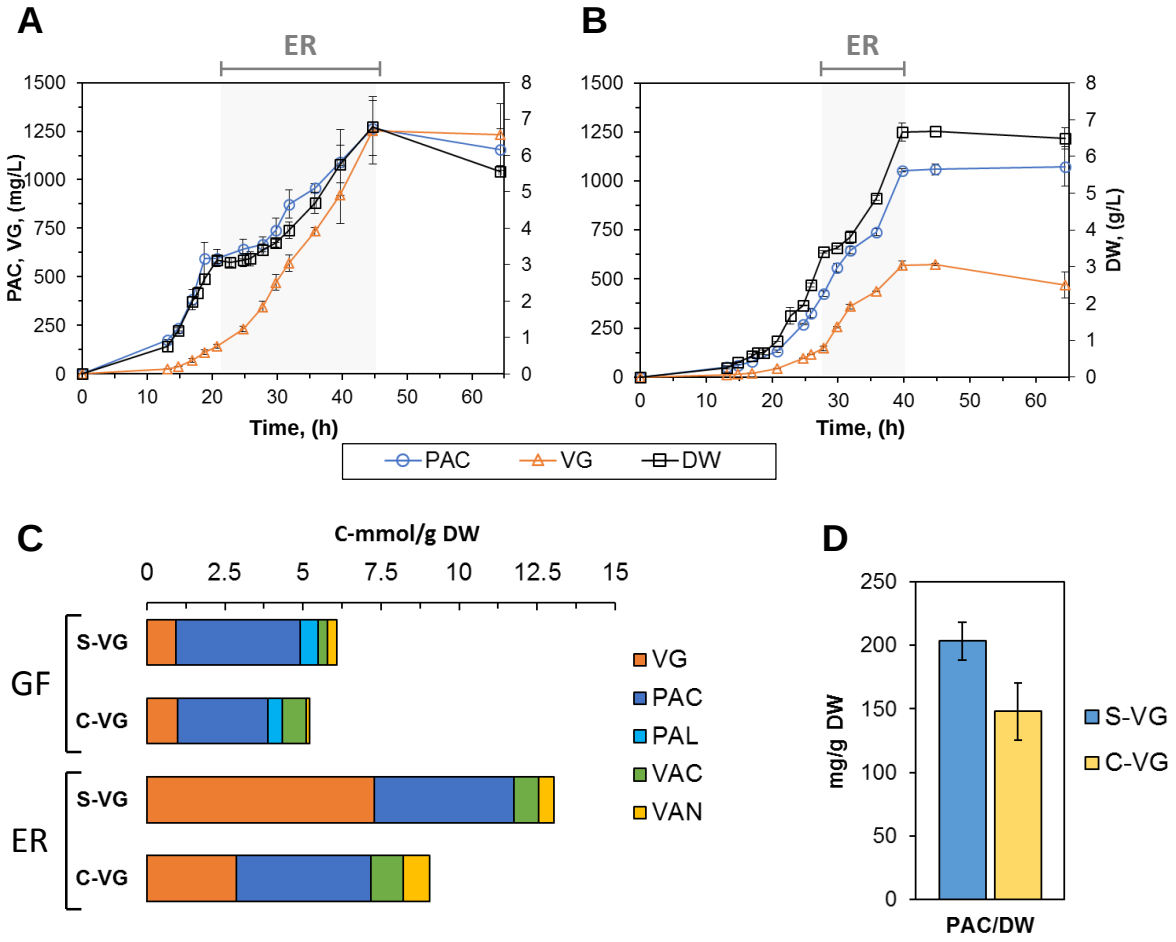


Figure 6.5. Distribution of VG and its intermediates in batch cultivation. Production profiles for biomas (DW) and metabolites PAC and VG in strains: A) S-VG and B) C-VG. Panel C) represents distribution of VG pathway metabolites at the end of different growth phases – glucose fermentation (GF) and ethanol respiration (ER). D) Average of PAC per DW rations calculated throughout all batch cultivation.

Metabolic profiles of VG producing strains in continuous cultures

In contrast to batch cultivation, HPLC analysis of the samples taken during continuous cultures revealed more prominent changes in the VG metabolic profiles of S-VG and C-VG strains (**Figure 6.6A** and **Table 6.5**). The activity of the VG pathway in the C-VG strain was decreased fourfold compared to S-VG, which can be seen from the total amount of carbon obtained by heterologous pathway (**Figure 6.6B**). Intermediates PAL and VAN were produced in trace quantities, the VAC accumulated to approx. 4 mg/g_{Glc} in both strains, while biosynthesis of the first intermediate PAC was pronounced in S-VG strain. Lastly, more than 10 fold higher VG yields based on glucose were observed in S-VG compared to C-VG, which well correlated with VG production rates (2.47 and 0.18 C-mmol/(g_{DW}·h), respectively). The fraction of the total carbon converted to VG was much higher in S-VG (62%) than in C-VG (17%) during chemostat cultivations.

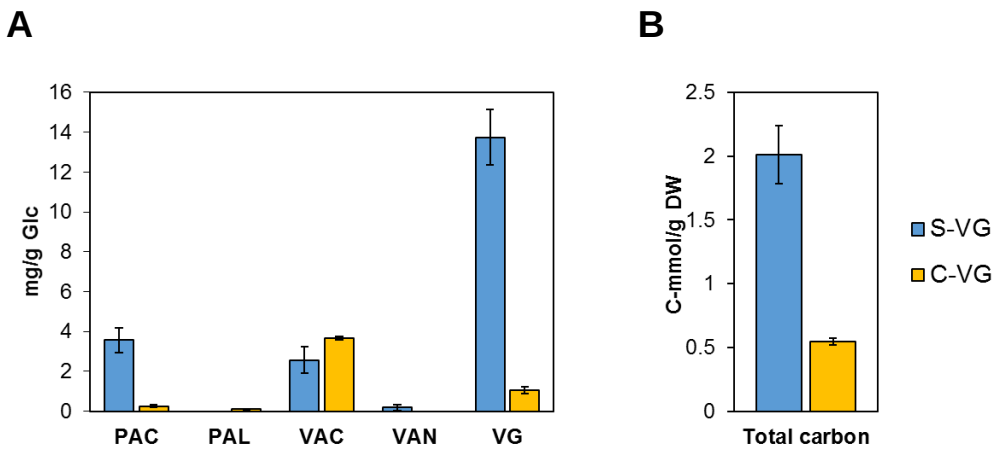


Figure 6.6. Final yields of VG and its intermediates produced by S-VG and C-VG during continuous cultivation; A) average of last three samples with more than one retention time in between, B) sum of all detected VG metabolites normalized to dry-weight.

Measuring intracellular cofactors associated with heterologous pathway enzymes

Since VG production is associated with several cofactors (**Figure 6.1**) we set out to measure their intracellular concentrations in two producing strains S-VG and C-VG. This was done to check if VG production differences might be a result of cofactor limitation. To ensure equivalent conditions for both strains, samples were taken during chemostat cultivations (see Materials and Methods). In particular, cofactors responsible for key enzymes (ACAR and UGT) in VG biosynthesis were considered. As displayed in **Table 6.6** concentrations of ATP, NADP and UDP-Glc were very similar in two VG producing strains and more importantly comparable to the ones obtained in several other studies were strains were cultivated under the same conditions [6], [33].

Table 6.6. Concentrations of internal metabolites during chemostat cultivations (µmol/ g DW).

Cofactor concentrations		
	S-VG	C-VG
ATP	8.51 ± 0.50	10.56 ± 1.25
ADP	0.91 ± 0.06	1.10 ± 0.06
NADP	0.35 ± 0.01	0.30 ± 0.01
NADPH	<0.02	<0.02
UDP	0.31 ± 0.02	0.32 ± 0.03
UDP-Glc	0.50 ± 0.02	0.53 ± 0.02
UTP	1.70 ± 0.20	1.45 ± 0.23

Discussion

The importance of the *Saccharomyces cerevisiae* in development of novel yeast cell factories is evident by the increasing number of research projects focusing on heterologous production of wide range industrially relevant substances [1], [34]. In the last decade, yeast strains S288c and CEN.PK or their derivatives are most frequently studied and used for cell factory development by the yeast community [2], [6]. In this study we demonstrated, however, that heterologous production of VG was significantly better in S288c based strain compared to CEN.PK based background. Assuming the transcription/translation efficiency in two strains is similar we have assured identical "genomic" conditions for integrated VG pathway in both strains by using well defined platform [11]. Therefore, based on the results of our study we speculate that the major changes in VG production were caused by the genotypic (consequently, phenotypic and/or physiologic) differences of S288c and CEN.PK strains [3], [6], [9], [10].

The heterologous production of VG in yeast is comprised of multi-enzymatic four step conversion with two possible biosynthetic routes (**Figure 6.1**).The existence of the two biosynthetic routes indicates wide substrate specificity of the heterologous enzymes, especially ACAR and HsOMT [12] as discussed previously [12]. The metabolic flux through the VG biosynthetic pathway is determined by 1) the availability the precursor 3-DHS, an intermediate of shikimate pathway, and 2) the activity of the heterologous enzymes and availability of cofactors associated with these enzymes. Since there are no cofactors associated with biosynthesis of first intermediate (PAC) [12] it was hypothesized that flux towards VG pathway should be dependent on the activity of shikimate pathway or, in other words, available pools of 3-DHS. The hypothesis was supported by strong correlation between PAC production and biomass yield discovered during batch cultivation in both S-VG and C-VG strains (**Figure 6.5D**). Moreover, the same correlation was previously detected by Brochado et al. [32] in the original VG producing (S288c based) strain.

In this experiment the comparison of S-VG and C-VG was troubled by the accumulation of intermediates of the VG pathway. The latter is a result of a non-optimally functioning heterologous enzyme in yeast cells. Despite the recent pathway optimization efforts [35], accumulation of intermediate metabolites remained unsolved. We speculate that unbalanced conversion to VG was mainly caused by low activities of heterologous enzymes and not by limitation of associated cofactors. In fact, intracellular concentrations of main cofactors such as ATP, NADPH and UDP-Glc (**Table 6.6**) were found to be similar in both S-VG and C-VG strains and moreover comparable to wild-type S288c and CEN.PK [6], [33]. Since the intermediate accumulation profiles were slightly different in two strains we therefore calculated the total metabolic flux through VG pathway for comparison purposes.

In batch, the total carbon produced by the heterologous pathway differed by 35% between the two strains, however, most of that difference was caused by a nearly twofold higher amount of VG produced by the S-VG compared to C-VG. The clear advantage of VG production in S-VG appeared only after the strain switched to the ER growth phase. It is evident from **Figure 6.5** that S-VG accumulated more of VG due to its nearly twice prolonged ethanol growth phase. The particular phenotype of S-VG is a result of the inherent deficiency in respiro-fermentative metabolism of ethanol found in the wild-type S288c strain [9]. Unfortunately, the reasons for impaired growth on ethanol in S288c when compared to CEN.PK could not be explained by neither transcriptional profiling nor genome sequencing efforts [9], [10]. The last enzymatic step in biosynthesis of VG (the UGT72E2 enzyme) itself is highly unlikely to play a significant role for the VG production difference between engineered strains. In this step vanillin is glycosylated by the UGT72E2 glycosyltransferase enzyme which uses UDP-Glucose as a donor for glycosylation [12]. Assuming that UGT was expressed identically in two strains there were no indications that its activity was restricted by the lower availability of UDP-Glc in C-VG strain as the volumetric productivity rates of VG in both strains were similar. In addition, in the previous (**Chapter 5**) it was shown that overexpression of UGT increased VG production in CEN.PK based strain.

As mentioned before, the role of alcoholic fermentation in VG biosynthesis in S-VG and C-VG strains is vague, however, the fact of plausibly different shikimate pathway activity in two strains cannot be ignored. In the batch cultivation experiments, metabolic flux through the VG pathway was found to be lower in the C-VG strain during both GF and ER growth phases (**Figure 6.5**). To investigate physiology in the absence of ethanol production, we have characterized S-VG and C-VG strains in chemostats with a standard cultivation conditions [6]. As we can see in **Figure 6.6B** metabolic flux through the VG pathway is 4-fold higher in S-VG compared to C-VG, and VG production more than 10-fold different **Figure 6.6A and Table 6.5**). Here we speculate that VG production is most likely restricted by decreased flux through the native shikimate pathway of C-VG and, therefore, less available 3-DHS pools. Recalling the correlation of PAC and biomass production observed in the batch experiment, the latter assumption is also supported by nearly absent accumulation of its direct metabolite PAC in chemostat cultures. Moreover, in a study by Canelas [6], the intracellular concentration of shikimic acid (the direct metabolite of 3-DHS) was found to be significantly higher in the YSBN (S288c based) strain under similar cultivation conditions in batch and chemostat cultivations. The shikimate pathway is a part of aromatic amino acid biosynthesis and is well characterized in yeast [36]–[38], furthermore, it is tightly regulated via two 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthase isoenzymes ARO3 and ARO4, which are feedback inhibited by phenylalanine and tyrosine, respectively **Figure 6.1**. Canelas and co-workers

have reported that under batch conditions CEN.PK113-7D is able to maintain higher intracellular amino-acid pools [6], which was suggested and later explained as a result of increased protein turnover rate [39]. However, here we speculate that higher concentrations of intracellular aromatic amino acids in C-VG might inhibit ARO3 and ARO4 and lead to decreased flux towards 3-DHS and consequently lower the capacity for VG production. Despite the fact that shikimate pathway is well understood in *S. cerevisiae*, there are no reported studies where this pathway is compared in two strains S288c and CEN.PK, and the exact cause of difference in shikimate pathway is not known. However, genome sequencing of CEN.PK113-7D revealed (considering only metabolic genes) 219 SNPs where 85 of those resulted in non-synonymous amino acid substitution [9]. The mutations were also found in genes encoding for proteins involved in amino acid and aromatic compound metabolism. Here, we have compared ORFs sequences of several ARO genes in strains S288c and CEN.PK (**Table 6.7**). For example, the gene of pentafunctional enzyme ARO1 (which is directly involved in synthesis of 3-DHS) is enriched with 15 SNPs, seven of those results in non-synonymous amino acid substitutions. In addition to that, the gene of the regulatory protein ARO3 was found to contain 23 SNPs, where three of those caused amino acid substitutions. However, phenotypic or metabolic repercussions of these mutations are unidentified, and need to be further characterized by targeted mutagenesis in one of the producing strains.

Table 6.7. Point mutations of ARO genes found in CEN.PK113-7D compared to S228c. Mutations represented by one letter code for amino acid and number denoting the position in the protein.

Genes	SNPs		Amino acid substitution
	Total	Non-synonymous	
ARO1	15	7	T225I, P337S, S517P, N844T, M1141K, V1386I, G1576A
ARO3	23	3	K141R, E214D, S349T
ARO4	—	—	—
ARO7	—	—	—

Yet another way to test the role of shikimate pathway in VG production is by alleviating feedback inhibition of key regulatory points of shikimate pathway in C-VG strain. This can be done by expressing inhibition insensitive ARO3 and ARO4 [38]. The improved VG production in C-VG with deregulated shikimate pathway would serve as strong evidence supporting the hypothesis of different aromatic amino acid metabolism in S288c and CEN.PK strains. One needs to add that production differences between strain backgrounds 288c and CEN.PK might be also true for other heterologous compounds which are using aromatic amino acids or intermediates of shikimate biosynthetic pathway as their precursors, for example indolylglucosinolates [11], flavonoid precursor - naringen [40] and other aromatics [41], [42].

The genotypic/phenotypic differences between two commonly used yeast strains [5], [6], [9], [10] may also be reflected in the production of other non-aromatic related heterologous compounds; e.g., antimalarial drug artemisinin precursors – artemisinic acid [43] or amorphadiene [44] are derived from ergosterol biosynthetic pathway which was found to be different in strains S288c and CEN.PK [9]. Particular differences in ergosterol pathway also led to rational target selection for metabolic engineering towards triterpene production in yeast [45].

Conclusions

In conclusion, we have demonstrated that the genetic background of two commonly used yeast strains had a significant impact on heterologous production of VG. The S-VG strain (s288c based) accumulated double the amount of VG in batch and more than 10 fold in chemostat cultures compared to the C-VG (CEN.PK based). The difference in VG production was defined by two major factors; first, in batch cultures, the production advantage in S-VG was an indirect result of inherent physiology of S288c (prolonged ethanol respiration phase), second, in batch and chemostat cultures VG production in the CEN.PK based strain was most probably limited due to reduced flux through yeast native shikimate pathway, which was also evident by significantly reduced accumulation of the first heterologous intermediate PAC. To this end, the comparison study for heterologous production of VG, an aromatic compound class metabolite, presented here might serve as useful information for correct yeast strain choice for future yeast based cell factory design. This study also indicated the importance of high quality systems biology datasets available for commonly used yeast strains which can be used in strain selection for building a cell factory.

Future perspectives

Several follow up studies for the particular project should be done. First, the role of shikimate pathway in different VG production in two strains should be tested by introducing point mutations in ARO3/4 which would alleviate feedback inhibition by tryptophan and phenylalanine [38]. By implementing aforementioned modifications in CEN.PK based strain we could easily check if increased flux through shikimate pathway in S-VG strain would result in higher production of VG. Secondly, more global approach can be implemented. Regardless of the fact that VG producing capacity was shown to be significantly better in S288c based strain, the CEN.PK series strains exhibit better physiological characteristic [3] and is more preferable strain for industrial applications. The VG production and physiological properties are defined by s288c and CEN.PK strain genetic backgrounds. One can speculate that combining these properties of two strains in one could result in super producing strain with advantageous physiological properties of CEN.PK. Unfortunately, exact gene sets responsible for particular behavior of both strains toward VG production are not

yet identified. This can be tested by creating hybrid strains, i.e., crossing S-VG and C-VG strains and screening a vast number of spores, which would contain a random set of genes from both strains, while the VG biosynthetic pathway would remain intact as it was engineered in an identical manner. If improved traits were discovered, the genomes of these mutants should be sequenced and analyzed. Obtained genetic set could explain not only the improved properties for the heterologous VG production, but give hints and clues for designing better yeasts with improved characteristics for cell factory development for production of other chemicals.

References

[1] I.-K. Kim, A. Roldão, V. Siewers, and J. Nielsen, “A systems-level approach for metabolic engineering of yeast cell factories.,” *FEMS Yeast Res.*, vol. 12, no. 2, pp. 228–48, Mar. 2012.

[2] K.-K. Hong and J. Nielsen, “Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries.,” *Cell. Mol. Life Sci.*, vol. 69, no. 16, pp. 2671–90, Aug. 2012.

[3] van Dijken JP, J. Bauer, L. Brambilla, P. Duboc, J. Francois, C. Gancedo, M. Giuseppin, J. Heijnen, M. Hoare, H. Lange, E. Madden, P. Niederberger, J. Nielsen, J. Parrou, T. Petit, D. Porro, M. Reuss, van Riel N, M. Rizzi, H. Steensma, C. Verrips, J. Vindeløv, and J. Pronk, “An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains.,” *Enzyme Microb. Technol.*, vol. 26, no. 9–10, pp. 706–714, Jun. 2000.

[4] K. Entian and P. Kötter, “Yeast genetic strain and plasmid collections,” *Methods Microbiol.*, vol. 36, no. 06, pp. 629–666, 2007.

[5] a Rogowska-Wrzesinska, P. M. Larsen, A. Blomberg, A. Görg, P. Roepstorff, J. Norbeck, and S. J. Fey, “Comparison of the proteomes of three yeast wild type strains: CEN.PK2, FY1679 and W303.,” *Comp. Funct. Genomics*, vol. 2, no. 4, pp. 207–25, Jan. 2001.

[6] A. B. Canelas, N. Harrison, A. Fazio, J. Zhang, J.-P. Pitkänen, J. van den Brink, B. M. Bakker, L. Bogner, J. Bouwman, J. I. Castrillo, A. Cankorur, P. Chumnanpuen, P. Daran-Lapujade, D. Dikicioglu, K. van Eunen, J. C. Ewald, J. J. Heijnen, B. Kirdar, I. Mattila, F. I. C. Mensonides, A. Niebel, M. Penttilä, J. T. Pronk, M. Reuss, L. Salusjärvi, U. Sauer, D. Sherman, M. Siemann-Herzberg, H. Westerhoff, J. de Winde, D. Petranovic, S. G. Oliver, C. T. Workman, N. Zamboni, and J. Nielsen, “Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains.,” *Nat. Commun.*, vol. 1, no. 9, p. 145, Jan. 2010.

[7] M. T. a P. Kresnowati, W. a van Winden, M. J. H. Almering, a ten Pierick, C. Ras, T. a Knijnenburg, P. Daran-Lapujade, J. T. Pronk, J. J. Heijnen, and J. M. Daran, “When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation.,” *Mol. Syst. Biol.*, vol. 2, p. 49, Jan. 2006.

[8] P. Daran-Lapujade, J. Daran, P. Kotter, T. Petit, M. Piper, and J. Pronk, “Comparative genotyping of the laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays,” *FEMS Yeast Res.*, vol. 4, no. 3, pp. 259–269, Dec. 2003.

[9] J. M. Otero, W. Vongsangnak, M. a Asadollahi, R. Olivares-Hernandes, J. Maury, L. Farinelli, L. Barlocher, M. Osterås, M. Schalk, A. Clark, and J. Nielsen, “Whole genome sequencing of Saccharomyces cerevisiae: from genotype to phenotype for improved metabolic engineering applications.,” *BMC Genomics*, vol. 11, no. 1, p. 723, Jan. 2010.

[10] J. F. Nijkamp, M. van den Broek, E. Datema, S. de Kok, L. Bosman, M. a Luttik, P. Daran-Lapujade, W. Vongsangnak, J. Nielsen, W. H. M. Heijne, P. Klaassen, C. J. Paddon, D. Platt, P. Kötter, R. C. van Ham, M. J. T. Reinders, J. T. Pronk, D. de Ridder, and J.-M. Daran, “*De novo* sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology.,” *Microb. Cell Fact.*, vol. 11, no. 1, p. 36, Jan. 2012.

[11] M. D. Mikkelsen, L. D. Buron, B. Salomonsen, C. E. Olsen, B. G. Hansen, U. H. Mortensen, and B. A. Halkier, “Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform.,” *Metab. Eng.*, vol. 14, no. 2, pp. 104–11, Mar. 2012.

[12] E. E. H. Hansen, B. B. L. Møller, G. R. G. Kock, C. M. Bünner, C. Kristensen, O. R. Jensen, F. T. Okkels, C. E. Olsen, M. S. Motawia, and J. Hansen, "*De novo* biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (Saccharomyces cerevisiae).," *Appl. Environ. Microbiol.*, vol. 75, no. 9, pp. 2765–74, May 2009.

[13] M. H. H. Nørholm, "A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering.," *BMC Biotechnol.*, vol. 10, p. 21, Jan. 2010.

[14] F. Geu-Flores, H. H. Nour-Eldin, M. T. Nielsen, and B. a Halkier, "USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products.," *Nucleic Acids Res.*, vol. 35, no. 7, p. e55, Jan. 2007.

[15] H. H. Nour-Eldin, B. G. Hansen, M. H. H. Nørholm, J. K. Jensen, and B. a Halkier, "Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments.," *Nucleic Acids Res.*, vol. 34, no. 18, p. e122, Jan. 2006.

[16] S. Partow, V. Siewers, S. Bjørn, J. Nielsen, and J. Maury, "Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae," *Yeast*, no. July, pp. 955–964, 2010.

[17] U. Güldener, S. Heck, T. Fielder, J. Beinhauer, and J. H. Hegemann, "A new efficient gene disruption cassette for repeated use in budding yeast.," *Nucleic Acids Res.*, vol. 24, no. 13, pp. 2519–24, Jul. 1996.

[18] S. Nakagawa, Y. Niimura, T. Gojobori, H. Tanaka, and K. Miura, "Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes.," *Nucleic Acids Res.*, vol. 36, no. 3, pp. 861–71, Feb. 2008.

[19] D. R. Cavener and S. C. Ray, "Eukaryotic start and stop translation sites.," *Nucleic Acids Res.*, vol. 19, no. 12, pp. 3185–92, Jun. 1991.

[20] D. Gietz, A. St Jean, R. a Woods, and R. H. Schiestl, "Improved method for high efficiency transformation of intact yeast cells.," *Nucleic Acids Res.*, vol. 20, no. 6, p. 1425, Mar. 1992.

[21] J. D. Boeke, F. Lacroute, and G. R. Fink, "Short communication A positive selection for mutants lacking orotidine-5 ' -phosphate decarboxylase activity in yeast : 5-fluoro-orotic acid resistance," *Mol. Gen. Genet.*, pp. 345–346, 1984.

[22] D. a Treco and F. Winston, "Growth and manipulation of yeast.," *Curr. Protoc. Mol. Biol.*, vol. Chapter 13, no. April, p. Unit 13.2, Apr. 2008.

[23] R. K. Mortimer and J. R. Johnston, "Genealogy of principal strains of the yeast genetic stock center.," *Genetics*, vol. 113, no. 1, pp. 35–43, May 1986.

[24] G. Bertani, "Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli.," *J. Bacteriol.*, vol. 62, no. 3, pp. 293–300, Sep. 1951.

[25] F. Sherman, G. R. Fink, and J. B. Hicks, *Laboratory course manual for methods in yeast genetics*. Cold Spring Harbor Laboratory, 1986.

[26] C. Verduyn, E. Postma, W. a Scheffers, and J. P. Van Dijken, "Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.," *Yeast*, vol. 8, no. 7, pp. 501–17, Jul. 1992.

[27] L. Olsson and J. Nielsen, "On-line and in situ monitoring of biomass in submerged cultivations," *Trends Biotechnol.*, vol. 15, no. December, pp. 517–522, 1997.

[28] S. G. Villas-Bôas, J. Højer-Pedersen, M. Akesson, J. Smedsgaard, and J. Nielsen, "Global metabolite analysis of yeast: evaluation of sample preparation methods.," *Yeast*, vol. 22, no. 14, pp. 1155–69, Oct. 2005.

[29] O. Magdenoska, J. Martinussen, J. Thykaer, and K. F. Nielsen, "Dispersive solid phase extraction combined with ion-pair ultra high-performance liquid chromatography tandem mass spectrometry for quantification of nucleotides in Lactococcus lactis.," *Anal. Biochem.*, vol. 440, no. 2, pp. 166–77, Sep. 2013.

[30] P. Van Hoek, "Effect of specific growth rate on fermentative capacity of baker's yeast," *Appl. Environ. …*, vol. 64, no. 11, pp. 4226–4233, 1998.

[31] H. C. Lange and J. J. Heijnen, "Statistical reconciliation of the elemental and molecular biomass composition of Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 75, no. 3, pp. 334–44, Nov. 2001.

[32] A. R. Brochado, C. Matos, B. L. Møller, J. Hansen, U. H. Mortensen, and K. R. Patil, "Improved vanillin production in baker's yeast through in silico design.," *Microb. Cell Fact.*, vol. 9, no. 1, p. 84, Jan. 2010.

[33] H. C. Lange, M. Eman, G. van Zuijlen, D. Visser, J. C. van Dam, J. Frank, M. J. de Mattos, and J. J. Heijnen, "Improved rapid sampling for in vivo kinetics of intracellular metabolites in Saccharomyces cerevisiae.," *Biotechnol. Bioeng.*, vol. 75, no. 4, pp. 406–15, Nov. 2001.

[34] E. Nevoigt, "Progress in metabolic engineering of Saccharomyces cerevisiae.," *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 3, pp. 379–412, Sep. 2008.

[35] A. R. Brochado and K. R. Patil, "Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering.," *Biotechnol. Bioeng.*, vol. 110, no. 2, pp. 656–9, Feb. 2013.

[36] S. Teshiba, R. Furter, P. Niederberger, G. Braus, G. Paravicini, and R. Hütter, "Cloning of the ARO3 gene of Saccharomyces cerevisiae and its regulation.," *Mol. Gen. Genet.*, vol. 205, no. 2, pp. 353–7, Nov. 1986.

[37] G. H. Braus, "Aromatic amino acid biosynthesis in the yeast Saccharomyces cerevisiae: a model system for the regulation of a eukaryotic biosynthetic pathway.," *Microbiol. Rev.*, vol. 55, no. 3, pp. 349–70, Sep. 1991.

[38] M. a H. Luttik, Z. Vuralhan, E. Suir, G. H. Braus, J. T. Pronk, and J. M. Daran, "Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: quantification of metabolic impact.," *Metab. Eng.*, vol. 10, no. 3–4, pp. 141–53, 2008.

[39] K.-K. Hong, J. Hou, S. Shoaie, J. Nielsen, and S. Bordel, "Dynamic 13C-labeling experiments prove important differences in protein turnover rate between two Saccharomyces cerevisiae strains.," *FEMS Yeast Res.*, vol. 12, no. 7, pp. 741–7, Nov. 2012.

[40] F. Koopman, J. Beekwilder, B. Crimi, A. van Houwelingen, R. D. Hall, D. Bosch, A. J. a van Maris, J. T. Pronk, and J.-M. Daran, "*De novo* production of the flavonoid naringenin in engineered Saccharomyces cerevisiae.," *Microb. Cell Fact.*, vol. 11, p. 155, Jan. 2012.

[41] J. O. Krömer, D. Nunez-Bernal, N. J. H. Averesch, J. Hampe, J. Varela, and C. Varela, "Production of aromatics in Saccharomyces cerevisiae--a feasibility study.," *J. Biotechnol.*, vol. 163, no. 2, pp. 184–93, Jan. 2013.

[42] L. Liu, H. Redden, and H. S. Alper, "Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces.," *Curr. Opin. Biotechnol.*, pp. 1–8, Mar. 2013.

[43] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. a Ho, R. a Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling, "Production of the antimalarial drug precursor artemisinic acid in engineered yeast.," *Nature*, vol. 440, no. 7086, pp. 940–3, Apr. 2006.

[44] P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman, and C. J. Paddon, "Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 3, pp. E111–8, Jan. 2012.

[45] K. M. Madsen, G. D. B. R. K. Udatha, S. Semba, J. M. Otero, P. Koetter, Y. Ebizuka, T. Kushiro, and G. Panagiotou, "Linking Genotype and Phenotype of Saccharomyces cerevisiae Strains Reveals Metabolic Engineering Targets and Leads to Triterpene Hyper-Producers," vol. 6, no. 3, 2011.

[46] S. R. Engel and J. M. Cherry, "The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database.," *Database (Oxford).*, vol. 2013, Jan. 2013.

Supplementary Information

Contents

List of Supplementary Tables

List of Supplementary Figures

Table S 6.1. List of verification primer pairs used for validation of gene targeting events. Appearance of the correct size band on the gel verifies targeted integration of your favorite fragment (XII-1, XII-2 and XII-5). In case of deletions, band size confirms your favorite gene deletion. *In case of ADH6 the band size also includes KanMX cassette.

Site	Name	Sequence	Fragment size, bp
XII-1	XII-1-up-out-sq	CTGGCAAGAGAACCACCAAT	729
XII-2	XII-2-up-out-sq	CGAAGAAGGCCTGCAATTC	739
XII-5	XII-5-up-out-sq	CCACCGAAGTTGATTTGCTT	688
—	C1_TADH1_F	CTTGAGTAACTCTTTCCTGTAGGTC	
For	*Deletion*	*Verification*	
BGL1	BGL1_VF	TCATCCTCCCTGTGTTTACA	1044
	BGL1_VR	AGTTGAAACAGAGGATAAGGTG	
*ADH6**	ADH6_VF	GTTTTGCTTTTTTCTCTGGG	2773
	ADH6_VR	GGAGTATCAACCACTAAAGCG	

Table S 6.2. The list of the additional primers used for sequencing purposes.

Name	Sequence
S1_TEF_F	CGGTCTTCAATTTCTCAAGTTTC
S1_PGK_F	CAAGGGGGTGGTTTAGTTTAGT
S1_ACAR_F	CATTGTTCGTTCCAGACACTGAC
S2_ACAR_F	GAAATTGTCTCAAGGTGAGTTCG
S3_ACAR_F	TTTGGGTAGATTCTTGTGTTTGG
YF19	AAAAAATAAATAGGGACCTAGACTTCA
YF21	GACCTACAGGAAAGAGTTACTCAAGAAT

Table S 6.3. List of the open reading frames (ORF) surrounding the integration sites used for VG pathway transfer.

Site	ORF	Function (from SGD[46])
XII-1 (upstream)	*RPS25B*	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S25, no bacterial homolog; RPS25B has a paralog, RPS25A, that arose from the whole genome duplication
XII-1 (downstream)	*YLR334C*	Dubious open reading frame unlikely to encode a protein, based on available experimental and comparative sequence data; overlaps a stand-alone long terminal repeat sequence whose presence indicates a retrotransposition event occurred here
XII-2 (upstream)	*SPO77*	Meiosis-specific protein of unknown function, required for spore wall formation during sporulation; dispensable for both nuclear divisions during meiosis
XII-2 (downstream)	*FKS1*	Catalytic subunit of 1,3-beta-D-glucan synthase; functionally redundant with alternate catalytic subunit Gsc2p; binds to regulatory subunit Rho1p; involved in cell wall synthesis and maintenance; localizes to sites of cell wall remodeling; FKS1 has a paralog, GSC2, that arose from the whole genome duplication
XII-5 (upstream)	*ILV5*	Bifunctional acetohydroxyacid reductoisomerase and mtDNA binding protein; involved in branched-chain amino acid biosynthesis and maintenance of wild-type mitochondrial DNA; found in mitochondrial nucleoids
XII- 5 (downstream)	*ATG33*	Mitochondrial mitophagy-specific protein; required primarily for mitophagy induced at post-log phase; not required for other types of selective autophagy or macroautophagy; conserved within fungi, but not in higher eukaryotes; ATG33 has a paralog, SCM4, that arose from the whole genome duplication

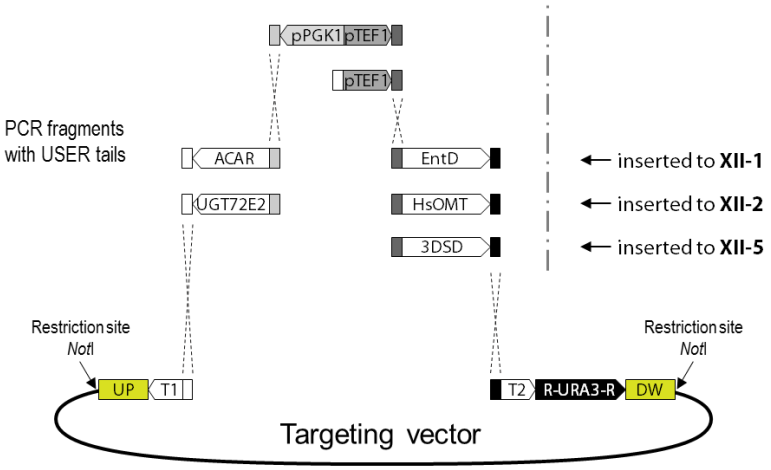


Figure S 6.1. Schematic representation of cloning procedure for transferring vanillin-β-glucoside pathway genes. Targeting vectors pXII1, pXII2 and pXII5 are linearized with AsiSI and Nb.BsmI enzymes. Yellow boxes UP and DW are targeting sequences homologous to chromosomal integration sites. T1 and T2 are terminator sequences Tadh1 and Tcyc1, respectively. Black arrow is counter-selectable auxotrophic marker URA3. White, grey and black boxes represent USER cloning compatible tails.

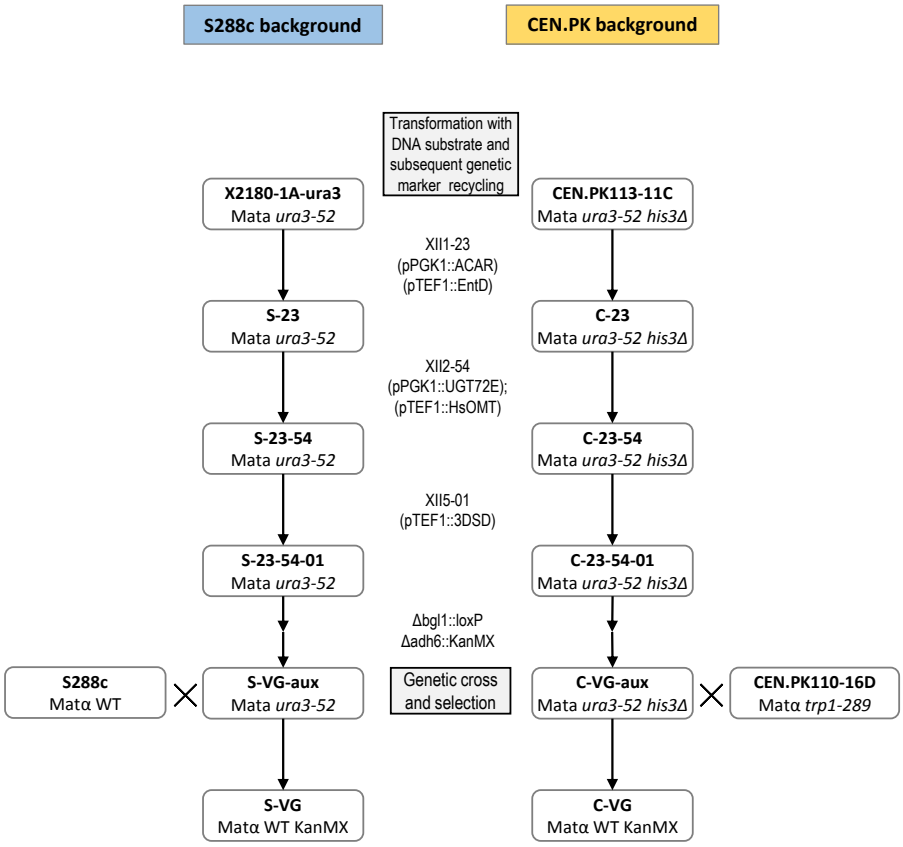


Figure S 6.2 Flowchart representing reconstruction of Vanillin-β-glucoside producing strains. *De novo* biosynthetic pathway was reconstructed in two strains by sequential transformation with appropriate DNA fragments depicted in the center of the figure. Native genes were deleted with *loxP*-KanMX-*loxP* cassette. Prototrophy was restored by genetic cross.

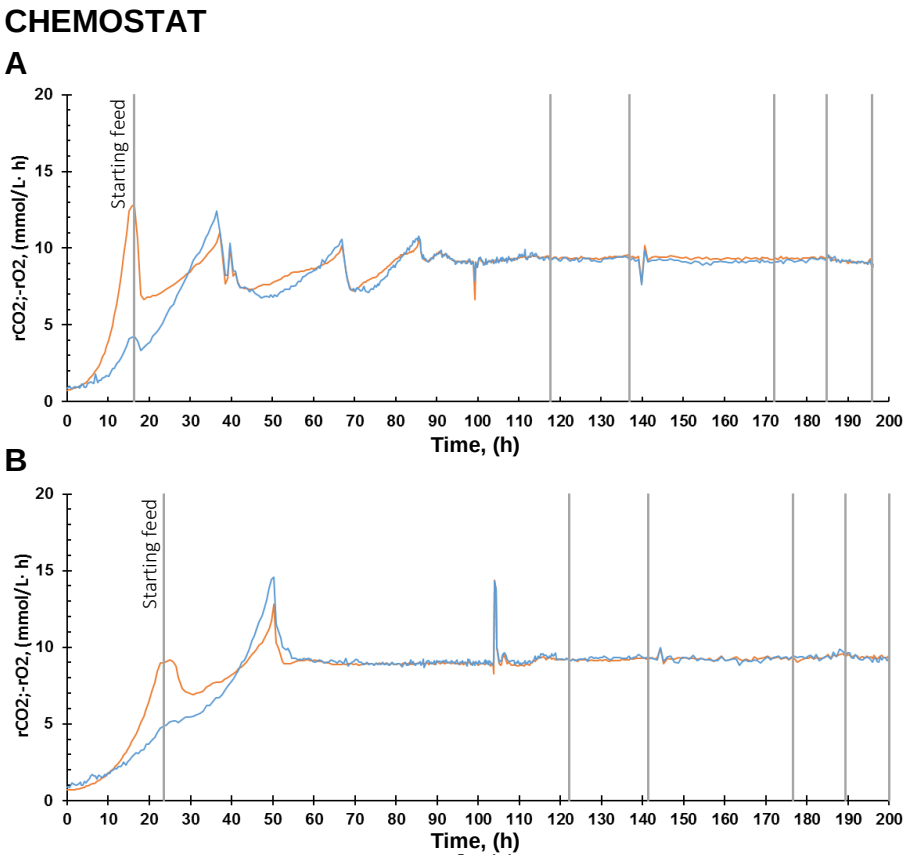


Figure S 6.3 Time profiles of online fermentation data of two strains A) S-VG and B) C-VG in chemostat cultivation mode. Orange line – volumetric rates of CO_2 production; red line – volumetric rates of oxygen consumption. Grey vertical lines represent the sampling points and the start of the chemostat phase. One residence time is equal to 10 hours.

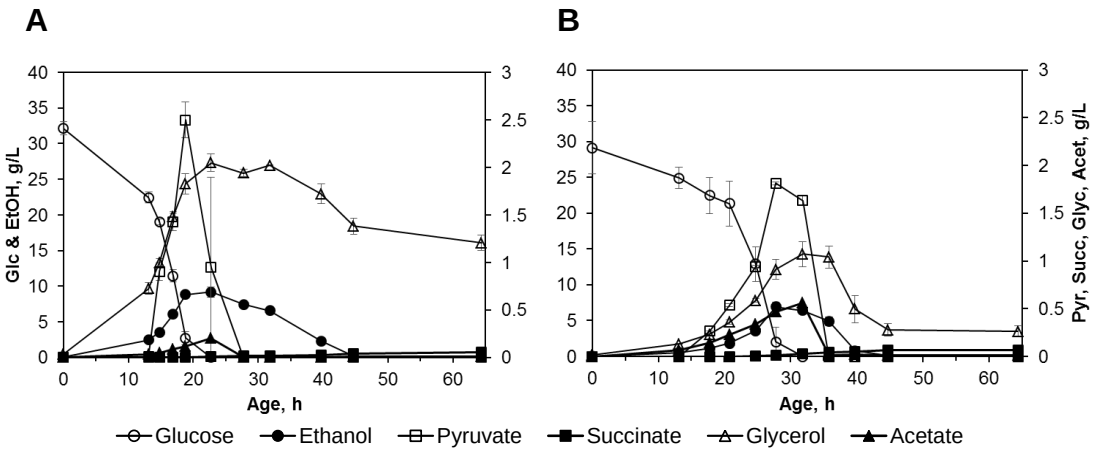


Figure S 6.4 Primary metabolite profiles in strains S-VG (a) and C-VG (b) during batch cultivation.

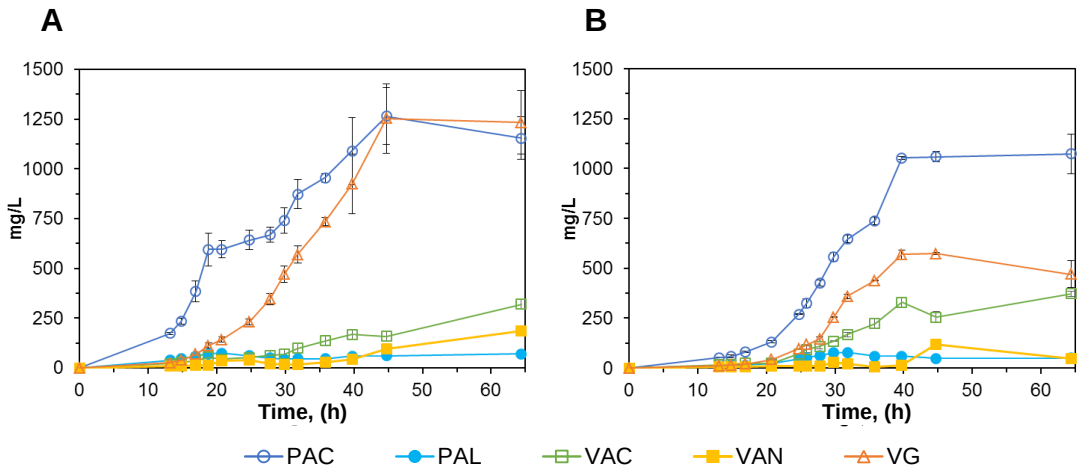


Figure S 6.5 VG metabolite profiles in strains S-VG (A) and C-VG (B) during batch cultivation.

Chapter 7

Conclusions and future perspectives

The work presented in my thesis dealt with the development and validation of genetic engineering tools that can be useful for yeast cell factory construction and metabolic engineering purposes. Two platforms have been successfully developed in this work: i) integrative vector set for rapid assembly and integration of large DNA fragments into the yeast *S. cerevisiae,* and ii) gene amplification system for controlled and stable gene overexpression. The latter system was further successfully applied for a real problem i.e., titration of unbalance metabolic pathway for vanillin-β-glucoside production. Moreover, an important point in yeast cell factory development, i.e., *S. cerevisiae* strain choice for the production of a given product (vanillin-β-glucoside), was addressed. In respect to this, it was demonstrated that two commonly used laboratory strains S288c and CEN.PK produced significantly different yields of vanillin-β-glucoside. Main conclusions and future perspective for each research project were outlined at the end of corresponding chapters (**Chapters 4** to **6**).This chapter provides with overall conclusions and general perspectives of the thesis.

Developed methods described in **Chapter 3** and **Chapter 4** were validated and proven to be efficient by expressing/overexpressing genes encoding for different fluorescent proteins. The EasyClone platform (**Chapter 3**) is an upgraded version of the integrative vector set previously developed at our center. It consists of 11 integrative vectors that allow integration of total of 22 genes of interest. Moreover, due to built-in safety system, it is possible to introduce repetitive DNA elements (the same promoter or gene can be used many times). The most appealing property of the EasyClone method is the short time required for gene assembly and integration. It was demonstrated that it is possible to introduce up to three integration cassettes (i.e., six genes) into *S. cerevisiae* genome in one transformation step. In this respect, the integration platform will be very useful for the construction of cell factories expressing large biosynthetic pathways in a high throughput manner. For example vanillin-β-glucoside pathway that consist of five genes could be integrated in one transformation step, in contrast, it took three rounds of transformation when using the first generation gene integration platform (**Chapter 2**).

The gene amplification method (**Chapter 4**) developed in this work was shown to be capable to amplify gene/s of interest up to ten copies. More importantly, it was demonstrated that strains

with overexpressed genes of interest were able to sustain stable expression levels in long term cultivations, which is a desired property for industrial applications. The fact gene amplification system allows overexpression of gene/s to a precise copy number and defined genomic location, makes the particular system very attractive for metabolic engineering applications. This was successfully demonstrated in **Chapter 5**, where gene amplification was used for balancing the heterologous production of vanillin-β-glucoside. The gene amplification system was designed to be expandable, by using specially designed targeting vectors the number of simultaneously amplified genes can be expanded from two to eight. In this case, an entire metabolic pathway can be overexpressed up to ten copies. Finally, the collection of gene amplification strains exist in various genetic markers and mating type combinations, which enables further manipulations via sexual cross (e.g., construction of diploid cells that was widely applied in vanillin-β-glucoside pathway titration experiments).

To this end, EasyClone and gene amplification system were designed to be compatible with high-throughput USER cloning method. Particular cloning method provides flexibility and enables combinatorial approaches when constructing expression cassettes. Therefore, genetic engineering tools developed in this work can be successfully applied for systemic biological studies to understanding production related processes of *S. cerevisiae.*

Future perspectives

The tool development for yeast genetic engineering is constantly evolving process. On the other hand, more attention has to be addressed to the genetic differences of the available *S. cerevisiae* strains. For example, genetic tools developed in my PhD project are based on CEN.PK background strain, which is mostly used for metabolic engineering studies mainly due to its industrially relevant phenotypic properties. However, the comparison study (**Chapter 6**) has indicated that S288c strain background is producing significantly higher yields of vanillin-β-glucoside than CEN.PK. This implies that the gene amplification system is not optimal for production of vanillin-β-glucoside and, most likely, other aromatic molecules. However, it might be opposite for the metabolites that are derived from other than shikimate pathway precursors. Moreover, these properties might be possible to predict from the available systems biology data sets for S288c and CEN.PK wild type strains.

Instead of using “pure” strain backgrounds, hybrid *S. cerevisiae* strains might be implemented. For example, to improve the gene amplification system, selected gene amplification strain (CEN.PK) can be crossed to a wild-type S288c strain. Subsequently, a vast number of spores would be screened to identify hybrid strain exhibiting desired properties, e.g., fast growth rate, high metabolic fluxes

through required pathways (e.g., shikimate pathway) and existence of gene amplification cassettes. If a haploid with desired traits would be obtained, the full genome sequencing could give hints and clues for development of "next generation" cell factories for the production of complex high-value natural and unnatural chemicals.